Annex A

Midterm Impact Analysis

NORTHERN GHANA MILLENNIUM VILLAGES IMPACT EVALUATION

Date: **November 2016**

Submitted by Itad

In association with:

External Impact Evaluation of the Millennium Villages Project, Northern Ghana

Midterm Impact Analysis

Table of Contents_Toc436731995

Table of Tables

Table of Figures

Acronyms and Abbreviations

AIDS	Acquired Immunodeficiency Syndrome
ANCOVA	Analysis of Covariance
CV	Control Village
DD	Difference-in-Difference
DHS	Demographic and Health Survey
FDR	False Discovery Ratio
GER	Gross Enrolment Ratio
GSS	Ghana Statistical Service
HIV	Human immunodeficiency virus
IPW	Inverse Probability Weights
IR	Imbens and Rubin
MDGs	Millennium Development Goals
MICS	Multiple Indicator Cluster Survey
MV	Millennium Village
NGO	Non-Governmental Organisation
NHIS	National Health Insurance Scheme
PPP	Purchasing Power Parity
PRG	Peer Review Group
PTA	Parent Teacher Association
WASH	Water, Sanitation and Hygiene
WHO	World Health Oragnisation

1. Introduction

This midterm report follows the analysis plan approved by the Peer Review Group (PRG). We start by discussing the quality of the data and its suitability for a difference-in-difference analysis. We then illustrate the impact of the Millennium Villages (MVs) on the Millennium Development Goals (MDGs) after two years of intervention without adjusting for differences in baseline characteristics. Next we present the methodology for the identification of programme effects. We then discuss in detail the impact of the intervention on poverty, income, food security, child health and education. Finally we discuss the heterogeneity of programme impacts across gender and administrative district, and we explore the presence of spatial spillover effects.

Data quality and suitability for difference in difference analysis

The quality of the expenditure and income data from previous rounds was analysed using Benford tests and was found to be comparable to similar budget surveys conducted in Ghana by the Ghanaian Statistical Office. We do not repeat the same analysis here though for some outcomes, such as child mortality, we will discuss issues of data quality in the appendices. The validity of a difference-in-difference approach however requires more than just data quality as it rests on the assumption that project and comparison groups are similar. Differential trends in the outcomes and covariates shocks were discussed in the previous second round analysis report and were not found to be major threats. Seasonality was found to be a serious threat for some outcomes and will be extensively discussed in the analytical section of this report in relation to the outcomes affected (anaemia, use of mosquito nets and reported morbidity). In this section we will focus on changes in the composition of the project and comparison groups produced by attrition, migration and measurement error. If attrition, migration and measurement error affect the two groups in different ways the difference-in-difference analysis can be biased.

1.1 Panel structure of the data

The baseline survey targeted a sample of 755 households in the MVs and 1,496 households in the Control Villages (CV). However, the baseline sample comprises only 711 MV households and 1,461 CV households because not all the target households were found at the time of the interviews. The second round of interviews and the midterm surveys targeted the same 755 MV households and 1,496 CV households originally selected at the baseline. The largest number of household interviews was conducted in the second round followed by the third round and the baseline round. There appear to be no systematic differences in the completion of survey interviews between project and comparison areas in any of the three rounds (Table 1). If anything, more households appear to be interviewed in CV areas than MV areas at the baseline, and the difference decreases over the three rounds (98% CV against 94% MV at baseline, 99% CV versus 98% MV in the second round and 97% in both CV and MV at the midterm).

Table 1. Completed household interviews during the first three rounds

Sample	Target	2012	2013	2014
MV interviews	755	711	743	735
%		*0.942*	*0.984*	*0.974*
CV interviews	1,496	1,461	1,487	1,456
%		*0.977*	*0.994*	*0.973*
ALL interviews	2,251	2,172	2,230	2,191
%		*0.965*	*0.991*	*0.973*

These very high completion rates are the result of efforts to locate households in repeated visits and of a very limited number of cases of household relocation or dissolution. The numbers of the latter cases are reported in Table 2 and are negligible in all survey rounds.

Table 2. Reasons for not completing the interviews

Reason	2012	2013	2014
No. competent household member at home	21	1	8
Entire household absent	22		11
Interview postponed	10		
Interview refused	1		
Partly completed			
Dwelling vacant or destroyed		4	2
Dwelling not found	19	9	13
Household has relocated		6	15
Household dissolved or deceased		1	6
Other	6		4*
ALL	79	21	59

Because not all target households are interviewed every year and because the households that are not interviewed differ from year to year, the number of full panel households that are interviewed every year decreases over time. The decrement is however very small (Table 3). The baseline survey interviewed 97% of target households, the second round retained 96% of the panel target households, while the mid-term survey retained 94%. The absolute numbers of households not interviewed are so small (14 households in MV areas and 51 households in the CV areas were lost at midterm in comparison to the baseline) that a comparative analysis of the characteristics of attriters and non-attriters in MV and CV areas is hardly feasible. In addition, these are not proper 'attriters' but simply households that could not be interviewed in all three survey rounds. The number of households permanently leaving the sample at some point is even smaller. The 'attrition' is very small and very similar across the project and comparison groups so that differential attrition does not appear to be a major threat for our data.

Table 3. Completed household interviews during the first three rounds

Sample	Target	2012	2013	2014
MV panel interviews	755	711	707	697
%		*0.942*	*0.936*	*0.923*
CV panel interviews	1,496	1,461	1,454	1,424
%		*0.977*	*0.972*	*0.952*
ALL panel interviews	2,251	2,172	2,161	2,121
%		*0.965*	*0.960*	*0.942*

Though the rate of household attrition is very small the attrition among individuals is quite large (Table 4). Only 90% of the individuals originally enumerated at the baseline were again enumerated after two years (91% in MV areas and CV areas). This however is only partly the result of migration, and is largely produced by natural processes (deaths and births) and changes in the composition of the households that will be the subject of the following section.

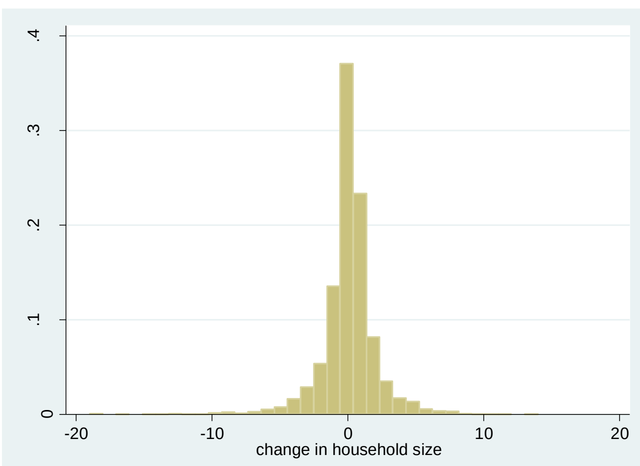
Table 4. Individuals listed in the surveys

Sample	2012	2013	2014
MV individuals	5,231	5,576	5,854
MV panel		*4,930*	*4,654*
CV individuals	10,337	10,649	11,023
CV panel		*9,869*	*9,378*
ALL individuals	15,568	16,225	16,877
ALL panel		*14,799*	*14,032*

1.2 Changes in household size and household composition

The sample population is relatively stable and slightly increasing over the surveys. This however conceals great changes in the composition of households. The overall household size is stable across survey periods, (7.1 at baseline, 7.0 at second round and 7.2 at midterm) but there is considerable change in household composition as shown in Figure 1. We calculated average household size of panel households at baseline and at midterm and the changes in size between the two periods. Less than 40% of households preserve the same sample size across a two-year period and there are dramatic changes by than 10 household members. These changes reflect a high level of individual mobility across time though are comparable to changes observed in very different socio-economic contexts. For example, Halliday (2005) finds that in El Salvador fewer than 50% of households experienced no change across two survey years.

Figure 1. Change in household size between baseline and midterm



Given the level of individual mobility across households, several anthropologists have questioned the validity of household survey in West Africa. Guyer (1981) observes that every report of household survey in West Africa discusses at some point the problem of defining household membership and of maintaining records of people with high mobility rates. Clearly our survey is no exception. Guyer (1981) and Hill's (1986) arguments against the collection of 'household' data can be summarised in the following way. First, West African households are polygamous and very large. As a result, household membership changes frequently as new members are added to the original nucleus or move away from it. In addition, there are always a number of family-related individuals who reside temporarily with the household units as foster children, as supported brothers of the household head or simply as visitors, which further complicate the task of defining households. Second, the large household size and the presence of independent individuals within the household (it is common in polygamous household for women to live independently and preserve strong ties with the family of origin) conjure against the ability of respondents to provide correct information, including simple demographics such as the number of children and their age.

Our surveys paid great attention to the collection of correct demographic data following international standards. Our surveys collect data on household members by making a list of all people living in the household excluding lodgers, guests and relatives living elsewhere. The enumerator's manual used in the training defines a household as a common decision-making unit occupying the same residence and normally eating from the same pot following United Nation (UN) recommendations (Glewwe 2000). Polygamous households formed are enumerated separately if they constitute separate household units. Finally, people who spend less than six months in the household over the 12 months preceding the survey are not considered household members unless they are infants or head of household.

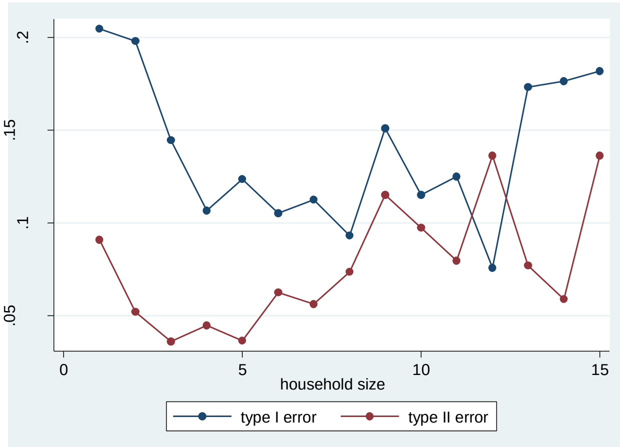
Of course good design and protocols do not automatically imply lack of errors. However, the collection of panel data allows the measurement of changes in household size across surveys and an estimation of reporting errors. Enumerators were instructed to re-compile the household roster at each round starting from the baseline roster, by adding new household members and removing those who were no longer members because of migration, death or error in reporting in previous enumeration. There is no guarantee that successive rounds of interview are free of error. However, our hope is that successive rounds identify errors committed in previous rounds by enumerators' probing and learning so that, in principle, the size of the error should decrease over time. To simplify, we define as type I error all individuals erroneously enumerated in previous rounds that should have not be enumerated, and we define as type II error all individuals enumerated in the current round that were erroneously not enumerated in previous rounds. The size of these errors across surveys is shown in Table 5. Type I error (individuals that were enumerated and should have not) is expressed in proportion of the enumerated individuals. Type II error (individuals who were not enumerated and should have) is expressed in proportion of the actual household members (the number of members enumerated minus those erroneously enumerated). Type I has decreased over time while type II error has remained stable. The error is 0 at the midterm by definition because we assume that the latest information collected on the ground is the correct one. Considering the high level of mobility in household composition, the errors are not large and always below 5%.

Table 5. Enumeration errors

Enumeration errors	Baseline	2nd round	Midterm
Household members enumerated	15,568	15,289	15,238
Actual household members	15,153	15,295	15,238
Type I error: non-household members (% of enumerated)	4.6	1.8	0.0
Type II error: missed household members (% of actual household members)	2.0	1.9	0.0

Errors are moderately correlated to household size. Figure 2 shows the fraction of respondents committing errors of the first and second type by household size. If respondents in large households have difficulties in remembering all household members the probability of making a type II error should increase with sample size. This is exactly what happens. Errors of the second type are more frequent among households of more than five members. Large households however do not appear to make errors of the first type, that is counting as members individuals who are not. This type of error is common to all households and more frequent among small households.

Figure 2. Probability of type I and type II errors at baseline



We further investigate the individual mobility across households by looking at those individuals who are considered household members by the respondent but who resided in the household for less than 12 months over the year preceding the interview. These individuals are normally removed by household surveys as 'non-household member' using some cut-off point. We calculate the fraction of this population after removing children who did not complete one year of age and type I error individuals. The results are reported in Table 6 including for household members who lived in the household for less than six months and their average age and sex composition. The percentage of individuals temporarily resident in the household is rather small and contrasts with the depiction of extremely fluid households by anthropologic studies. Movements of people between households do not appear too large and frequent.

Table 6. Non-permanent resident household members

Non-permanent household members	Baseline	2nd round	Midterm
% less than 12 months	5.0	2.9	3.6
% less than 6 months	1.6	1.2	1.2
% males (<6)	49.4	42.9	33.5
Average age (<6)	22.0	20.2	19.4

In general, population changes can only occur through the balancing population equation (Deaton 1997). There are only two ways for people to enter a population: being born or migrating into it. Similarly, there are only two ways to leave a population: death and out-migration. The population at any time is therefore the results in changes in the natural population growth (births minus deaths) between t and t-1, and changes in net migration (in-migrants minus out-migrants) between t and t-1, to which we add reporting errors: a reduction in population resulting from type I errors in the previous survey and an increase resulting from type II errors:

$$N_t = N_{t-1} + \left(B_{t-1,t} - D_{t-1,t}\right) + \left(I_{t-1,t} - O_{t-1,t}\right) + (e2_{t-1,t} - e1_{t-1,t})$$

The natural population change is simply the change determined by the difference between births and deaths. The MV intervention has an ambiguous impact on this rate as long as it decreases the number of deaths (by public health measures) but also the number of births (directly by family planning and indirectly by increases in wealth that lead families to invest in children quality rather than quantity). The potential impact of MV on migration should be less ambiguous. Note that migration is not meant here simply as movement outside a geographic area but as a movement outside or inside the household. Individuals moving out might, for example, be forming new households in the same village. The increase in income and productive opportunity brought about by the project should reduce migration outside the geographic area and at the same time increase household size through additional marriages, child fostering and incorporation of individuals related to households located in MV areas. A positive relationship between household size and wealth in polygamous societies has been theorised by Becker (1991) and observed empirically by Grossbard (1980) and Whitehead (2006). In general we may expect household size to increase (or decrease more slowly) in MV areas compared to CV areas as a result of migration.

The decomposition of population change in natural growth, migration and errors is shown in Table 7 for MV and CV areas separately using only those sample households (2,212) that were interviewed for three consecutive rounds. Rates of change are calculated over the number of household members in the previous round. The overall natural population change is positive and is very similar in MV and CV areas. The fraction of people moving in and out of the household is considerable but the two terms tend to cancel out and the changes are very similar in MV and CV areas. Enumeration errors also cancel out at the second round but not at the midterm when type I error is larger (thus leading to an increase in population as more people are found at the midterm that had not been previously and erroneously enumerated). Again the patterns of enumeration errors are very similar in MV and CV areas.

Table 7. Population changes across surveys

Population change	All		MV		CV	
	2nd round	Midterm	2nd round	Midterm	2nd round	Midterm
Overall change	-1.3	3.5	-1.3	2.1	-1.3	3.7
Natural change (births minus deaths)	0.8	1.5	0.9	1.5	0.7	1.5
Migration: people moving in	1.7	4.0	2.3	4.3	1.5	3.9
Migration: people moving out	-3.2	-4.0	-4.2	-4.5	-2.7	-3.8
Type I error	4.1	3.6	5.0	2.4	3.6	3.6
Type II error	-4.3	-1.5	-5.1	-1.9	-3.9	-2.1
Residual unexplained difference	0.4	-0.1	-0.2	0.4	-0.5	0.6

Overall the analysis conducted in this section shows that, despite great challenges, the surveys were able to enumerate households with great accuracy and that no large differences emerge between MV and CV areas in population changes driven by natural change, movements in and out of households and reporting errors. A simple comparison of household size across surveys and comparison groups confirms this observation. The average difference in household size is very small and never statistically significant at each survey round (baseline difference 0.20, P-value 0.481; 2nd round difference 0.20, P-value 0.395; midterm difference 0.17, P-value 0.551).

2. Participation in project activities

The MV projects offers a wide range of services in agriculture, health and education and makes considerable efforts in trying to reach all households. The services offered by MV however are also offered to some extent by other agencies in the same area. In this section we analyse the coverage of MV activities and whether their reach varies with poverty levels of the beneficiaries. Tables 8 through 11 show the participation rates in MV and CV areas in activities promoted by the project in the areas of: social mobilisation, agriculture, health and education. We summarise the main results as follows:

- Participation in groups and in the activities promoted by the MVP is high in control areas as well, where the same activities are promoted by the government or by other projects and non-governmental organisations (NGOs).
- MV does not have a great impact on social mobilisation (Table 8), with the exception of farmer-based organisations. There are a larger number of women's groups and parent teacher associations (PTAs) in MV areas than in CV areas, but the difference is not statistically significant. Participation in a wide range of social groups promoted by MV (including water, sanitation and hygiene (WASH) groups, MDG school groups, water and sanitation groups, mother-to-mother support groups, daddy's clubs, village savings and loans associations, and school management committees) is very small in both MV and CV areas and no differences are visible.
- There is a large and statistically significant impact on participation in agricultural extension, fertiliser use and access to loans in MV areas (Table 9).
- There is also a large and statistically significant impact on National Health Insurance Scheme (NHIS) membership and households' visits by health workers and related activities, such as provision of condoms, anthropometric measurements and advice on breastfeeding, child feeding and use of bednets (Table 10). On the other hand, clinic usage and related activities, such as supplementation of de-worming and vitamin A, are not larger in MV areas compared to CV areas.
- Children in MV areas have greater access to school feeding but there are no signs of greater access to other education benefits such as stationery, books and bursaries (Table 11).

Table 8. Membership of project-related groups

Social mobilisation groups	MV	CV	P-value	Observations
Cooperative	21.1***	5.2	0.000	2,191
Farmer-based organisation	23.9***	5.0	0.001	2,191
Farmer field school	0.8	0.4	0.398	2,191
Women's group	19.2	15.1	0.146	2,191
Parent-teacher association	41.0	35.0	0.218	2,191
WASH	0.9**	0.1	0.011	2,191
MDG school club	-	-	-	2,191
Water and sanitation development board	1.0	0.2	0.116	2,191
Mother-to-mother support group	0.5	0.1	0.238	2,191
Daddy's club	0.0	0.1	0.323	2,191
Village savings and loan association	3.0	2.5	0.725	2,191
School management committee	0.8	0.7	0.781	2,191

Table 9. Participation in agriculture-related activities

	MV	CV	P-value	Observations
Adult received agricultural training	47.8***	16.4	0.000	5,078
Any household member received a loan	11.8***	1.2	0.000	2,191
Used any fertiliser	55.2***	36.8	0.000	2,158

Table 10. Participation in health-related activities

Participation	MV	CV	P-value	Observations
Membership of NHIS	87.0***	55.5	0.000	15,958
Someone distributed bednets	32.4	36.6	0.452	2,191
Visit by a community health worker (CHW)	65.3***	20.5	0.000	2,191
The CHW provided condoms	21.6***	3.4	0.000	2,191
The CHW measured children's arms	48.3***	13.9	0.000	2,191
CHW advised on breastfeeding	50.9***	14.6	0.000	2,191
CHW advised on child feeding	55.5***	16.6	0.000	2,191
CHW advised on use of bednets	54.3***	17.4	0.000	2,191
Visited a health facility	75.0	71.4	0.303	2,191
Children given de-worming	40.3	35.2	0.124	2,191
Children given vitamin A	40.4	38.9	0.704	2,191
Children given food supplements	9.3***	1.8	0.000	2,191
Children given sanitary pads	2.0**	1.1	0.006	2,191

Table 11. Participation in education-related activities

Participation	MV	CV	P-value	Observations
Child had a school meal on previous day	43.7**	27.3	0.027	2,191
Children received a bursary	0.8	0.3	0.104	2,191
Children received stationery, uniform, etc.	24.6	20.6	0.356	2,191

Next we investigate whether the MV interventions are targeted or accessed to a different extent by beneficiaries of different economic status. We do so by plotting participation rates in project activities in MV and CV areas against baseline per capita expenditure. The goal of this exercise is to show the ability of the programme to reach the poorest sectors of the population. In Table 12 we report the results of a series of statistical tests that assess the statistical significance of the visual differences observed in the charts of Figure 3. The main results are as follows:

- The charts show no selective targeting of households by expenditure levels as the lines are flat over the whole expenditure distribution in most cases. This is confirmed by the null results of Test 2 in Table 12, which shows no pattern in MV participation by 5 expenditure quintiles. The only exception is agricultural extension which appears to increase slightly with income.

- There are no obvious patterns by per capita expenditure in participation into activities in the control group either (Test 1 in Table 12).

- As a result of the two points above, there are no differences in patterns between MV and CV with the exception of agricultural extension (less pro-poor in MV compared to CV areas) and fertiliser use (more pro-poor in MV compared to CV areas – upper quintiles of the distribution benefit less in MV).

- Overall there is no indication that the intervention targets the poorest of the poor. Rather the programme appears to be directed to the whole population, which in turn has equal access to the interventions. The poorest of the poor are not missed out.

Figure 3. Participation in project activities by household expenditure

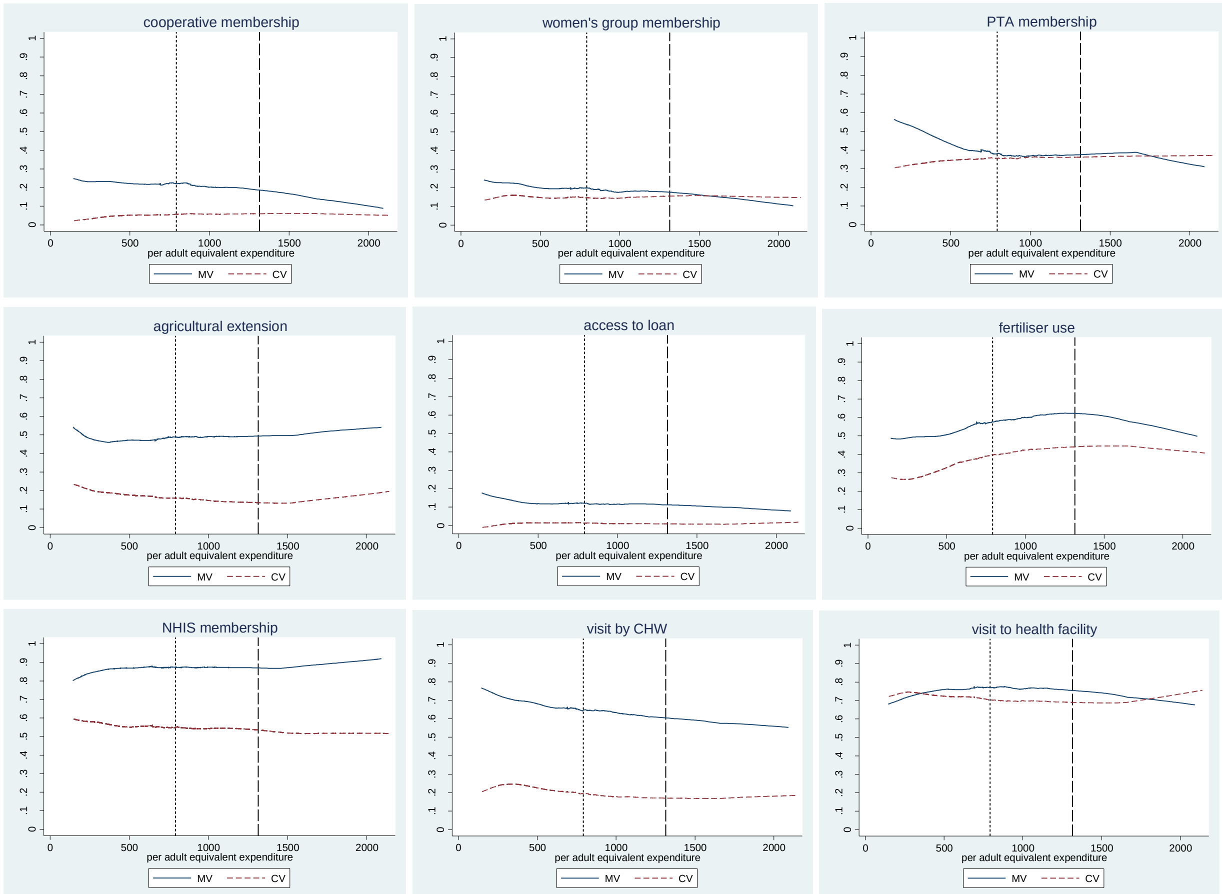


Table 12. F-tests of equality in participation patterns

Equity in participation	Test 1	Test 2	Test 3	Test 4
Membership of cooperatives	0.47	0.39	20.93***	0.34
Membership of women's groups	0.50	0.87	1.19	0.43
Membership of PTA	0.09	0.91	2.21	1.23
Agricultural Extension	1.45	3.05**	36.17***	2.80**
Loans	1.23	0.52	9.16***	1.25
Fertiliser use	1.47	1.66	3.11**	2.15**
Membership of NHIS	1.29	0.32	10.26***	0.51
Visits by community health workers	0.84	0.39	20.93***	0.34
Visits to clinics	0.59	0.27	1.26	1.05

Tests are conducted by running regressions of participation dummies on quintiles of baseline per-adult equivalent expenditure of MV and CV households. Test 1 evaluates the presence of a pattern in CV areas (equality of all CV slope coefficients). Test 2 evaluates the presence of a pattern in MV areas (equality of all MV slope coefficients). Test 3 evaluates the difference between MV and CV areas at each quintile (all MV slope coefficients are zero). Test 4 evaluates the difference in patterns between MV and CV areas (joint difference between MV and CV slope coefficient).

3. Impact of MV on the Millennium Development Goals

Table 13 shows the impact of the intervention on the MDGs at the midterm. Impacts are expressed as difference-in-differences: before-after changes in the outcomes in the MV areas minus the same changes in the comparator group. All variables considered are dichotomous and the outcomes are reported as ratios, shares, proportions or rates. Details on the calculation of each indicator using the household survey data are reported in Appendix A1. In some cases our calculation of the MDG indicator differs slightly from the official MDG definition but care was taken in reproducing the official methodology exactly. P-values are reported in parentheses and are obtained from statistical tests through regressions of the outcomes on time and project variables corrected for clustering. In some cases the regression approach was not feasible and a bootstrapping of the standard errors at the cluster level was employed (p-values obtained by bootstrapping are indicated in the table). All impacts are calculated from cross-sectional data without exploiting fixed-effects and without adjusting for baseline differences in characteristics between the MV and CV groups. Since many hypotheses are tested at the same time, there is a possibility that some rejections are the result of chance. There is a chance that some impacts are found statistically significant when they are not. We first discuss the size of the impacts and their statistical significance using standard approaches and we then move to discuss what impacts hold after adjusting for multiple testing. Smiley and un-smiley faces in the third column show statistically significant positive and negative project impacts, respectively.

Table 13. Impact of MV on the Millennium Development Goals

MDG	Difference-in-difference	DD impact (at 10% sig.)	FDR (at 10% sig.)	Bonferroni (at 10% sig.)
Goal 1 Eradicate extreme poverty and hunger				
Proportion of population below $1 (PPP) per day	-7.4 (0.132)			
Proportion of population below the national poverty line	2.3 (0.521)			
Poverty gap ratio	-3.9 (0.213)			
Share of poorest quintile in national consumption	1.7 (0.447)[a]			
Employment to population ratio	5.5* (0.088)	☺		
Proportion of employed people living below $1 (PPP) per day	-7.5* (0.095)	☺		
Proportion of own account and contributing family workers in total employment	4.9** (0.004)	☺	☺	☺
Percentage of underweight children under-5	0.8 (0.803)			
Proportion of population below minimum level of dietary energy consumption	-7.8 (0.159)			
Goal 2 Achieve universal primary education				
Net enrolment ratio in primary education	3.3 (0.282)			
Proportion of pupils starting grade 1 who reach last grade of primary	-3.2 (0.304)			
Literacy rate of 15-24 year-olds, women and men	-0.6 (0.674)			
Goal 3 Promote gender equality and empower women				
Ratio of girls to boys in primary education	-0.09			

	(0.292)[a]			
Ratio of girls to boys in secondary education	0.15 (0.860)[a]			
Ratio of girls to boys in tertiary education	-0.63 (0.761)[a]			
Share of women in wage employment in the non-agricultural sector	-24.2** (0.020)	☺		
Goal 4 Reduce child mortality				
Under-5 mortality rate	-1.12 (0.514)[a]			
Infant mortality rate	-0.09 (0.953)[a]			
Proportion of 1-year-old children immunised against measles	0.6 (0.890)			
Goal 5 Improve maternal health				
Proportion of births attended by skilled health personnel	16.7** (0.002)	☺	☺	☺
Contraceptive prevalence rate	4.1** (0.030)	☺		
Antenatal care coverage	-4.2 (0.127)			
Proportion of population aged 15-24 with comprehensive correct knowledge about HIV	0.1 (0.953)			
Proportion of children under-five sleeping under insecticides treated bednets	33.0*** (0.000)	☺	☺	☺
Goal 7 Ensure environmental sustainability				
Proportion of the population using an improved drinking water source	-7.6** (0.042)	☹		
Proportion of the population using an improved sanitation facility	0.7 (0.754)			
Goal 8 Develop a global partnership for development				
Fixed telephone subscriptions for 100 inhabitants	0.0 (0.997)			
Mobile cellular subscriptions for 100 inhabitants	-7.5 (0.201)			

All figures are percentages. P-values in parentheses.[a] P-values calculated using 1,000 bootstrap sample replications at the cluster level. In all other cases p-values are obtained from cross-sectional regressions of the outcome (0/1) on project dummies and using cluster standard errors.

MDG 1 is *Eradicate extreme poverty and hunger*. There is little impact on poverty though a larger impact is visible on extreme poverty (proportion of population below $1 (PPP) per day and proportion of population below minimum). The negative impact on the poverty gap ratio and positive impact on the share of expenditure going to the bottom quintile of the population seem to confirm that the intervention is helping the extremely poor. None of these impacts however is statistically significant. There are no differences in the prevalence of underweight among children. The intervention increases the proportion of individuals performing any work as shown by changes in the employment to population ratio and in the proportion of own account workers in total employment (the two ratios represent almost the same phenomenon because the predominant activity in the area is subsistence agriculture). Poverty among the employed has decreased as shown by the reduction in the proportion of the employed living below $1 PPP. The impacts observed on employment indicators are statistically significant.

The second MDG is *Achieve universal primary education*. No impacts are found in relation to this goal. Observed effects on net attendance of primary school, the proportion of children completing primary school and in literacy rates of adults aged 15-24 are very small and never statistically significant.

The third MDG is *Promote gender equality and empower women*. More girls than boys are attending school at all three education levels (primary, JHS and SSS) before the intervention in both MV and CV areas. The project has no impact on this ratio. There is a considerable reduction in the share of women in wage employment in the non-agricultural sector in the MV area. However, the proportion of the population employed in the non-agricultural sector is very small so that the absolute impact on employment of this change is very small.

MDG 4 is *Reduce child mortality*. The project caused a reduction in infant and under-5 mortality by about 10 per thousand points. This impact is calculated over a five-year interval, which heavily underestimate the actual impact because the programme has been running for only two years. The differences are not statistically significant, partly because of the relatively small sample size employed. There is no impact on the percentage of children immunised against measles.

Goal number 5 is *Improve maternal health*. The project has a large and statistically significant impact on the number of births attended by a skilled professional, on the use of contraceptives and on the proportion of children sleeping under insecticised mosquito bednets. The programme has no impact on antenatal care and on HIV knowledge.

MDG 7 is *Ensure environmental sustainability*. The programme has no impact on the proportion of households using improved toilet facility, while the impact on the proportion of households having access to drinking water appears to be negative. The effect is large and statistically significant, implying that access to safe water sources has worsened as a result of MV.

MDG 8 is *Develop a global partnership for development*. The project has not increased the use of mobile phones. This impact is negative and not statistically significant. No household owns landline phones in the area and no impact is observed.

To summarise, the data suggest that there was a drop in extreme poverty and a visible increase in employment in the MV areas. The data are also suggestive of a reduction in child mortality and there is a visible increase in births attended by skilled personnel, the use of contraceptive methods and of mosquito bednets. No impact of the project is found on: undernutrition, school attendance, girls' school attendance, measles vaccination rates, antenatal care, use of safe toilets and use of mobile phones. Finally, the data seem to suggest that access to safe drinking water has worsened as a result of the intervention.

This analysis is affected by the multiple testing problem. By setting a significance level of 10% in testing 28 hypotheses we are implicitly allowing for a 95% change of rejecting at least one null erroneously (1-0.90^28). This is a well-known statistical problem without an obvious solution.1 One popular approach to address the multiple testing problem is the use of Bonferroni-type corrections. Bonferroni corrections change the P-values in proportion to the number of hypotheses tested. A typical approach consists of dividing the statistical significance level by the number of tests (0.1/28=0.036 in our case) and using this level to test the significance of the differences found. The results of adopting this approach are reported

[1] For a review of approaches to multiple testing offered in the statistical literature, see the review by Schochet (2008).

in the last column of the table. Only three null hypotheses are rejected implying that only three project impacts are real: the increase of own account and contributing family workers in total employment; the increase in the proportion of births attended by skilled health personnel; and the increase in the proportion of children sleeping under insecticide-treated bednets.

Bonferroni-type corrections are considered too restrictive because they reduce the probability of type I errors, as intended, only by increasing the probability of not finding any impact when in fact there is one. A problem which is compounded when the outcome variables tested are highly correlated as in our case. Therefore, we also employ the false discovery rate approach (FDR), which calculates the fraction of null hypotheses that are wrongly rejected (Fink, McConnell and Vollmer 2014). Interestingly, in our application, the two approaches deliver exactly the same results. In both cases the number of statistically significant effects is reduced to just three.

Overall the results reported in Table 14 suggests that the impact of the intervention after two years is very limited. This analysis however is affected by a number of limitations:

- **Selection bias.** Impacts are estimated as if the observations had been obtained from a randomised experiment. These differences are not adjusted for baseline differences in characteristics between project and comparator areas.

- **Seasonality.** Baseline data were collected at different times in MV and CV areas. Previous analysis of the data have shown that some variables are unaffected by seasonal bias while others are. Among the MDGs reported, the following indicators are slightly affected by seasonal reporting: child nutrition and school attendance. One indicator, use of bednets, is greatly affected by seasonal bias as the use of nets is more common in the rainy season than in the dry season. This suggests that part of the large project impact observed on the use of bednets could be a consequence of this bias.

- **Sample sizes.** Some of the MDG indicators are calculated using small groups because they require samples constructed for specific age-sex categories. Small sample sizes result in large standard errors and a reduction in the likelihood of finding statistically significant small effects. Other indicators are based on large samples but because the results are small, the standard errors are nevertheless large: this is particularly the case of child mortality rates. Absence of statistical significance for these variables may reflect a lack of impact as well as a small sample size.

- **Inadequacy of old MDGs indicators.** The official MDG indicators do not describe accurately socio-economic conditions in the study area and are unable to detect the impact of the project on livings standards. For example, employment-related indicators have little relevance in a poor area where most individuals do some work at any time. On the other hand, relevant indicators of well-being like prevalence of anaemia and malaria or cognitive skills and education test scores are missing. In addition, MDG indicators based on prevalence rates pay little attention to changes in the distribution of the outcomes so that, for example, a reduction in the percentage of extremely malnourished children cannot be observed.

The analysis conducted in the following section of this report will address these limitations. In particular, comparisons will be made for non-MDG relevant welfare indicators and all comparisons will be performed using a combination of difference-in-difference analysis and matching methods, thereby adjusting for baseline difference in characteristics between the MV and CV groups.

4. Methodology

Project impact is estimated using difference-in-difference (DD) analysis: the difference in the change over time in the average outcomes between the project and in the comparator groups. In the simple standard two-period and two-group set-up, the difference-in-difference effect is:

$$\delta = (\bar{y}_{P,1} - \bar{y}_{P,0}) - (\bar{y}_{C,1} - \bar{y}_{C,0})$$

where δ is the DD effect, y is the average outcome either in the project group (P) or in the comparison group (C) observed in the first period (0) and in the second period (1).

We calculate DD effect using regression analysis. We use different regression models depending on whether (a) data are available for panel observations and (b) data are available for two or three periods. The simplest model is the **cross-sectional regression**:

$$y_i = a + bT_i + cP_i + dP_iT_i + \sum_{j=0}^{n} g_jX_i + e_i$$

where y is the outcome for the observation i, T is a dummy variable equal to 0 for period 1 and equal to 1 for period 2, P is a dummy variable equal to 1 if the observation is in the project group and equal to 0 if the observation is in the control group, PT is equal to 1 if the observation is both in the project group and observed in the second period. The equation estimates the following: a is the average outcome in the control group in period 1; b is the difference in the outcomes between period 2 and period 1 in the control group (the time trend); c is the difference between project group and control group in period 1; d is the DD effect of the project. The (X_i) are covariates that improve the balance between the project and comparison groups' samples, as these were not randomly obtained from an experiment, and increases precision of the estimates by reducing the standard error of the coefficients. One potential problem with the use of covariates in the estimation of project effects is that most covariates are affected by the project or are themselves objectives of the intervention. Think, for example, of a difference-in-difference (DD) regression of height-for-age including changes in total household expenditure. The inclusion of variables affected by the programme will 'absorb' some of the project effects that would otherwise be captured by project dummies. Hence, in order to capture the programme impact with a project dummy interaction, the covariates can only include variables that are not affected by the programme and can include baseline values of the variables Rosenbaum (1984).

When panel data are available we use a **fixed effects model** to remove the impact of fixed effects: time-invariant unobservable determinants of the outcomes such as, for example, farmers' motivation or children innate abilities. The fixed effect model is:

$$y_{it} = a_i + bT_i + dP_iT_i + \sum_{j=0}^{n} g_jX_{it} + e_i$$

The covariates in this case are time-varying variables that are not affected by the project such as, for example, the occurrence of drought or other shocks. As recommended by Angrist and Pischke (2009) we also employ the **lagged outcome model** (Imbens and Wooldridge 2009) also known as the analysis of covariance (ANCOVA) model:

$$y_{i1} = a + by_{i0} + dP_iT_i + \sum_{j=0}^{n} g_jX_{i0} + e_i$$

which is simply a regression of the dependent variable in period 2 on the dependent variable in period 1 and a project dummy in addition to the usual baseline covariates.

As an aside we note that these models can be expanded to include multiple time periods and for completeness we report below the model specifications employing three periods. For each of the three models above we report the specification estimating the average project effect over the three-year period and the specifications estimating two year-specific project effects.

Three-period cross sectional models:

$$y_{it} = a + b_1 T_{i1} + b_2 T_{i2} + cP_i + dP_i T_{it} + \sum_{j=0}^{n} g_j X_i + e_i$$

$$y_{it} = a + b_1 T_{i1} + b_2 T_{i2} + cP_i + dP_i T_{i1} + dP_i T_{i2} + \sum_{j=0}^{n} g_j X_i + e_i$$

Three-period fixed effects models:

$$y_{it} = a_i + b_1 T_{i1} + b_2 T_{i2} + dP_i T_i + \sum_{j=0}^{n} g_j X_{it} + e_i$$

$$y_{it} = a_i + b_1 T_{i1} + b_2 T_{i2} + d_1 P_i T_{i1} + d_2 P_i T_{i2} + \sum_{j=0}^{n} g_j X_{it} + e_i$$

Three-period lagged models:

$$y_{it} = a + b y_{it-1} + dP_i T_i + \sum_{j=0}^{n} g_j X_{i0} + e_i$$

$$y_{it} = a + b y_{it-1} + d_1 P_i T_{i1} + d_2 P_i T_{i2} + \sum_{j=0}^{n} g_j X_{i0} + e_i$$

The comparator villages surveyed at the baseline were identified by matching district villages to project villages using a propensity score built using village-level characteristics obtained from census data and from field visits. In order to remove remaining baseline differences in characteristics between the project and the control group we further employ matching methods at the household and individual level in the estimation of the project effects. In doing so we follow the methodology for the estimation of treatment effects under unconfoundedness outlined by Imbens and Rubin (2015).

Imbens and Rubin (IR) recommend the separation of the design stage from the analysis stage in conducting observational studies. The goal of the 'design' stage is selecting a propensity score and a sample of observations that maximises the statistical balance of the distribution of the covariates. In the design stage the outcomes are completely ignored in order not to bias the construction of the propensity score. The goal of the analysis stage is estimating project effects in the selected sample using the propensity score estimated in the design stage. We briefly describe the various steps followed in the design and analysis stage. In order to provide an example, all the results of the procedure adopted to estimate the impact of the project on per-adult equivalent expenditure are reported in Appendix A1.

Design stage:

- We estimate the propensity score using a logistic regression model. IR propose an algorithm for the estimation of the propensity score which aims at achieving statistical balance of the covariates and does not try to 'explain' participation through a behavioural model. The initial covariates are selected based on substantive knowledge of the existing literature and of the context. Covariates are subdivided in:
    - *Basic* covariates that are known to be strong determinants of the outcomes or of participation in the project. These covariates are included in the model regardless of the statistical significance of their correlation with project status.
    - Additional covariates that are likely to be correlated with the outcomes or with participation in the project. These covariates are added to the logistic model stepwise based on the statistical significance of their correlation with project status.
    - A group of powers and interactions of all the variables identified in groups (a) and (b) after the estimation of the model.
- We assess the validity of the estimated propensity score by testing the balance of the covariates. We first subdivide the sample based on the propensity score using the algorithm suggested by IR. We then conduct the three tests of sections 17.3.1, 17.3.2 and 17.3.3 of IR and we plot the figures of section 13.8.
- We assess the degree of overlap in the distribution of the covariates in the project and the comparator groups. We need to avoid that comparisons are made for observations, either in the project or comparison group, which have very few or no similar observations in the other group. To do this we inspect the distribution of the propensity scores using histograms and calculate the proportion of project and control observations with sufficiently 'good' matches using the method described in 14.5 of IR. We then trim the sample to remove the observations that are outside the region of overlap. The region of overlap is identified using the algorithm described in section 16.4 of IR.

Analysis stage:

- We first repeat some of the operations conducted at the design stage in order to calculate the propensity score and identify the size of the estimation sample. We run a logit participation model and we calculate the propensity score. We trim the sample to remove observation outside the region of overlap. We re-estimate the propensity score on the trimmed sample.
- We estimate project effects using inverse probability weights calculated using propensity scores and normalised following the procedure in 17.8.1 of IR. IR do not recommend inverse probability weights because (a) the estimated propensity score – used to build the weights – can be noisy and biased; (b) the propensity score may have a large variance because of extreme values close to zero and one; (c) the estimator uses fixed parameters while these are allowed to vary when using more flexible methods

like sub-classification. At this stage we conducted most estimations using inverse probability weights because a) we have to estimate impacts over a large number of outcomes and sub-classification is more laborious while the application of inverse probability weights is straightforward; b) in some cases (child mortality rates in particular) when the number of observed events is very small the reduction of the sample to small blocks and the calculation of effects within blocks is hardly feasible.

- In some instances we employ the sub-classification method recommended by IR and we will employ this method throughout in the next iteration of the report. When using sub-classification, we split the sample in blocks based on the propensity score until the statistical difference in the propensity score within each block is removed following the algorithm of section 13.5 of IR. We then calculate project effects as the weighted average of the block level treatment effects, where the weights are calculated using the proportions of control and project observations in each block. The methodology follows sections 17.3 and 17.5 of IR. Treatment effects within each block are difference-in-difference regressions.

5. Impact of the MVP on monetary poverty

We calculated poverty rates using the methodology and the poverty lines used by the Ghanaian Statistical Service (details of the methodology employed are reported in Appendix 3). Poverty rates for the first three survey rounds are reported in Tables 14 and 15 for MV and CV areas, respectively. The details on the calculation of these poverty rates can be found in Appendix A2. Poverty rates calculated in relation to a minimum basket of food and non-food items (the 'general poverty line') are reported in Table 14, while Table 15 shows poverty rates calculated in relation to the ability to purchase a minimum basket of food items (the 'food poverty line'). The first poverty rate is an indicator of overall poverty while the second is an indicator of extreme poverty. We also include two distributional measures of poverty: (a) the poverty gap, which shows the extent to which the poor are far from the poverty line; and (b) the squared poverty gap, which takes into account the distribution of expenditure among the poor, an indicator that gets larger when people at the bottom of the expenditure distribution get poorer.

All poverty indices are very similar in MV and CV areas at the baseline and at the second round. Overall poverty has not decreased at the midterm in MV areas. However, food poverty shows a considerable decrease at the midterm and all distributional poverty rates have reduced, pointing to a reduction in inequality.

Table 14. Poverty indices (overall poverty)

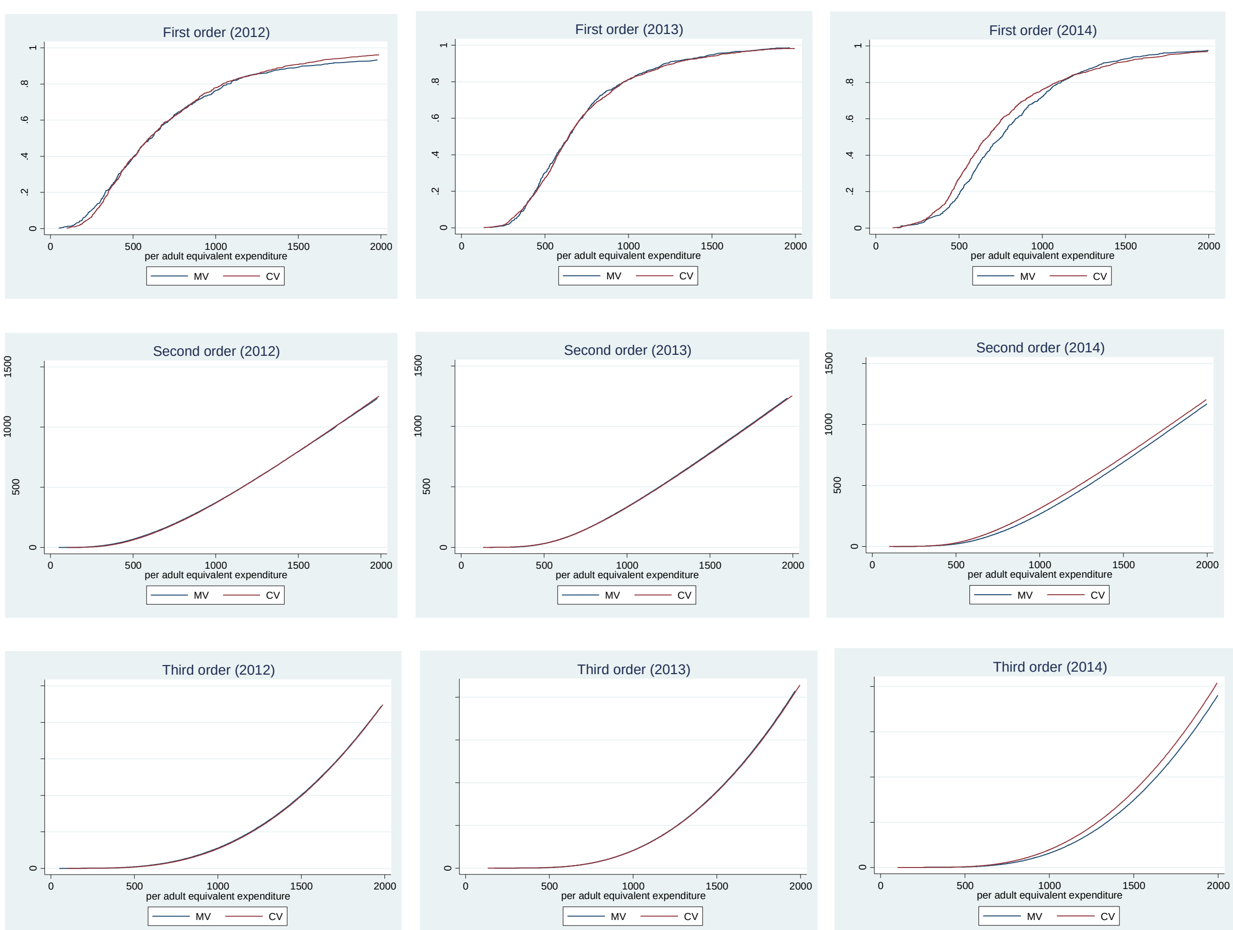
	Poverty headcount		Poverty gap		Squared poverty gap	
	MV	CV	MV	CV	MV	CV
Baseline	0.863	0.874	0.480	0.478	0.309	0.303
2nd round	0.915	0.911	0.462	0.458	0.266	0.264
Midterm	0.883	0.871	0.397	0.434	0.214	0.250

Table 15. Poverty indices (food poverty)

	Poverty headcount		Poverty gap		Squared poverty gap	
	MV	CV	MV	CV	MV	CV
2011/2012	0.652	0.647	0.283	0.275	0.156	0.145
2012/2013	0.686	0.672	0.223	0.222	0.095	0.095
2013/2014	0.548	0.620	0.166	0.210	0.070	0.093

Poverty rates are very high because the new poverty line introduced by Ghana Statistical Service (GSS) is very high, and most households in the sample are poor based on official poverty definitions. The question arises whether the poverty cut-off used in official statistics is appropriate for our sample of households. Is poverty robust to different poverty lines? In general, poverty rates are sensitive to the poverty line chosen, for example, because the project benefits some groups but not others at different positions in the expenditure distribution. The data suggest that at midterm overall poverty is higher in MV areas but that food poverty is lower. The robustness of poverty indicators to different poverty lines can be investigated by using 'poverty incidence curves' (Deaton 1997). Poverty incidence curves plot the proportion of poor people in the population at different levels of the poverty line. If the incidence curve for MV areas is above the CV curve along all the range of the expenditure distribution, then there is said to be first-order stochastic dominance and any poverty line can be used. If one curve is above the other only over a given range, then poverty lines around that range are appropriate but poverty lines near the crossing point are not. Stochastic dominance is analysed in the charts of Figure 4. The poverty headcount fails the stochastic dominance test. The curves of MV and CV areas are crossing so that poverty is larger over some range and lower over some other range. Curves of higher order, referring to poverty gap and squared poverty gap indices, do not fail the dominance test and the corresponding indicators are more robust and can be safely used.

Figure 4. Stochastic dominance of poverty indices



The lack of robustness of official poverty lines for our application suggests that the impact of the intervention would be better assessed by looking at changes in average per-adult consumption rather than looking at arbitrary welfare cut-offs and that the impact should be assessed along the expenditure distribution rather than estimating a single average effect for all households in the sample.

Impact of MV on per-adult equivalent expenditure

Here we estimate the average treatment effect of the project on per-adult equivalent expenditure. The outcome variable is the log of per-adult equivalent expenditure so that the regression coefficients of project dummies have a simple interpretation of percentage increase in expenditure brought about by the intervention. The results are reported in Table 16. We employ three different model specifications: simple cross-sectional analysis, a fixed-effect model and a lagged dependent variable model. All models are adjusted by baseline covariates correlated with per capita expenditure and by time varying covariates that are unrelated to the intervention such as household size and the occurrence of droughts and floods. The observations are weighted based on a propensity score that balances the baseline distribution of covariates of the project and comparator groups.

Table 16. Difference-in-difference impact on per-adult equivalent expenditure (IPW method)

	Cross-section	Fixed effects	Lagged model
Average DD effect	0.031	0.029	0.024
	(0.618)	(0.644)	(0.569)
DD effect second year	-0.005	-0.007	-0.009
	(0.938)	(0.922)	(0.829)
DD effect third year	0.069	0.066	0.057
	(0.324)	(0.347)	(0.337)
Sample size	5942	5942	3942

P-values in parentheses based on cluster adjusted standard errors.

Programme impacts are slightly smaller when they are adjusted by the probability of being in the sample (the propensity score). Overall the results show a very limited impact of the intervention on per-adult equivalent expenditure. The effect size is an average 3% expenditure increase in real terms per year, which disaggregates into a 0% impact in the first year and an impact of about 6.5% in the second year (we take the effect between the fixed effect model and the lagged model as the best approximation of the true effect). These effects are never statistically significant, though the p-values of the midterm coefficient are not too large. Our sample was not designed to detect statistical significance of extremely small effects such as those reported in Table 16, which could explain the absence of statistically significant effects. We also estimate programme impact using the sub-classification method proposed by Imbens and Rubin. The results are reported in Table 17. Both coefficient estimates and probability values are remarkably similar to those obtained using inverse probability weights (IPW).

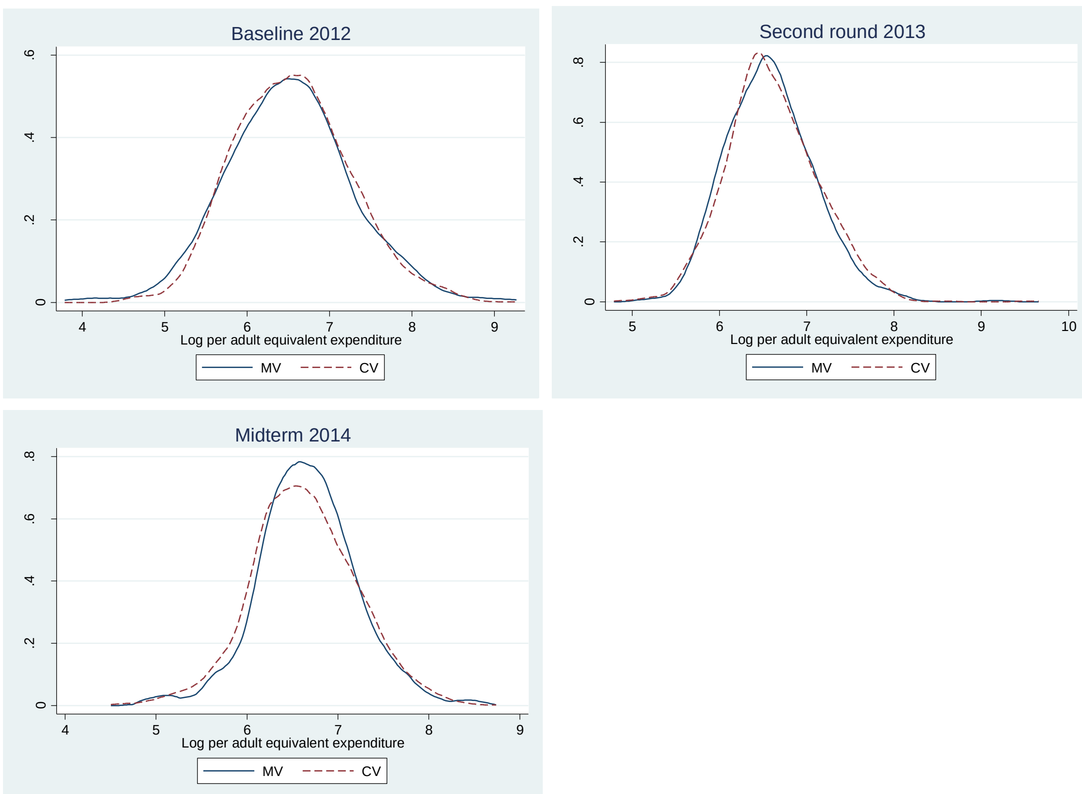
Table 17. Difference in difference impact on per-adult equivalent expenditure (sub-classification method)

	Fixed effects	Lagged model
Average DD effect	0.029	0.021
	(0.642)	(0.610)
DD effect second year	-0.009	-0.011
	(0.887)	(0.791)
DD effect third year	0.067	0.055
	(0.340)	(0.370)
Sample size	5942	3942

Distributional impact of MV on household expenditure

The distributions of per-adult equivalent expenditure over the three rounds in MV and CV areas are plotted in Figure 5. The distributions are nearly identical at baseline and at the second round, but there is a shift to the right of the expenditure distribution in the MV areas at the midterm. MV households are getting richer. This does not translate into an overall reduction of poverty as measured by the official poverty line because most changes occur below the poverty line.

Figure 5. Densities of log per-adult equivalent expenditure in the project and control groups



Further, we look at the impact of the intervention on the standard deviation of the logarithm of per-adult equivalent expenditure, a standard measure of inequality (Cowell 2011). This measure can be easily converted into another familiar inequality indicator: the Gini coefficient.[2] Both indicators are reported in Table 18. The distribution of per-adult equivalence expenditure has become more equal in the MV areas over time and on a difference-in-difference basis it reduced by 0.09 points, corresponding to a change of the Gini coefficient of 0.05.

[2] Assuming the distribution of per-adult equivalent expenditure is log-normal, the Gini coefficient is $2\Phi\left(\frac{\sigma}{\sqrt{2}}\right) - 1$, where sigma is the log of per-adult equivalent expenditure.

Table 18. Inequality of the distribution

	Standard deviation		P-value	Gini equivalent	
	MV	CV		MV	CV
Difference at baseline	0.739	0.700	0.292	0.40	0.38
Difference at 1st round	0.506	0.521	0.596	0.28	0.29
Difference at midterm	0.533	0.581	0.245	0.29	0.32
Difference-in-difference	-0.087*		0.053	-0.05	

*** is statistical significance at 1%, ** is 5% significance and * is 10% significance.

We then assess the impact of the intervention for households at different levels of the expenditure distribution using quantile regressions. We employ the lagged model specification for these regressions because it is the one with more statistical power to detect small effects and we calculated clustered standard errors by bootstrapping. The results are reported in Table 19. The estimated coefficients show that the impact of the intervention decreases with household per-adult equivalent expenditure. The impact of MV is stronger on the poor and very poor.

Table 19. Distributional impact on per-adult equivalent expenditure

	10th percentile	25th percentile	50th percentile	75th percentile	90th percentile
Average DD effect	0.038	0.008	0.008	-0.001	-0.003
	(0.493)	(0.864)	(0.877)	(0.987)	(0.966)
DD effect second year	-0.009	-0.028	-0.010	-0.029	-0.014
	(0.861)	(0.623)	(0.843)	(0.594)	(0.818)
DD effect third year	0.137	0.068	0.088	0.015	-0.007
	(0.179)	(0.416)	(0.266)	(0.830)	(0.916)

Difference-in-difference coefficients of lagged models. P-values in parentheses are corrected by clustering at the village level using bootstrap sample replications.

6. Impact of MVP on household income and food security

In this section we look at the impact of MV on household income. Income is calculated from the household questionnaire and can be separated into: agricultural income (farm profits calculated as the difference between the value of agricultural output minus production costs); livestock income (consisting of variation in stocks and value of production minus costs); employment incomes (income obtained from being employed in an economic activity in exchange for payment in cash or kind); business income (resulting from a self-employment activity and consisting of the difference between revenues and costs); and transfer incomes (consisting of remittances and other transfers in cash or kind, mostly related to development projects).

The calculation of income raises a number of difficulties. In particular, underreporting is likely resulting in an underestimation of household income. This is evident by comparing income figures to expenditure figures. Total household incomes are on average less than 50% of reported household expenditure. This is possible only if households are disinvesting assets or receiving transfers. However, transfers and livestock sales are already included in income computations so that some form of underreporting is occurring. Overestimation of costs might also explain part of the discrepancy between incomes and expenditures. At times, costs are so high in relation to outputs that incomes become negative. This is the case of between 10–15% of households in each round. Of course, negative incomes are a possible result of crop or livestock losses. However, in order to allow a minimum level of expenditure, they should be compensated by transfers or asset sales, which are also reported in the survey and in the computation of income. Finally, negative incomes and discrepancies with expenditures could be funded by loans, but the size of indebtedness reported by households is relatively small.

Table 20. Income shares by source, all households

Income source	Baseline	1st round	Midterm
Agriculture	56.0	45.1	42.1
Livestock	28.4	30.1	32.2
Business	7.1	16.5	16.9
Employment	4.5	4.3	5.5
Transfers	0.8	0.8	0.8

Though the absolute incomes of this sample of households are likely to be underestimated it is still worth analysing their composition and their changes over time. Composition and changes are not necessarily biased by underreporting. Table 20 shows that agriculture is the main income source for these households. The combination of income from agricultural production and livestock makes up about 80% of household income. Small business activities, such as petty trade and services, are the second most important income source. Few individuals are employed either casually or under longer contracts and the size of transfers, including remittances, is negligible. Of course, these figures refer to reported income and a different distribution of income sources may emerge if all incomes were reported correctly. For example, a larger share of income might result from petty trade carried out by different household members, or from remittances of temporary migrants if these are systematically underreported. There is a noticeable change in the composition of income sources over time. The share of agricultural income has decreased at the same time as livestock and business income have increased. It is difficult to interpret these patterns. They might be the result of agricultural shocks related to rainfall patterns, the impact of MV on different economic activities, or an improvement in enumeration.

We estimate the impact of MV on household income using the same approach employed in the estimation of the impact on household expenditure. Since a large fraction of incomes are negative we are not allowed to use logarithms. We decided to standardise the income figures by the baseline common standard deviation of income. The latter was approximately 1.5 times average income at baseline, so that changes in standard deviations can be translated to changes with respect to the mean with little effort. As before, we use three different difference-in-difference regression models: simple cross-section estimates, fixed effects and lagged dependent variable models. The estimation employs the same inverse probability weights used in the expenditure models, and the difference-in-difference regressions include determinants of income that may differ across the MV and CV groups.

The results are rather surprising (Table 21). Even the most conservative estimates of the lagged model suggest an increase in household income by 0.2 standard deviations corresponding to an increase of about 30% per year and all the effects are highly statistically significant. These results contrast with the small effects observed on household per-adult equivalent expenditure. At least two possible explanations are in order for the simultaneous large impact on income and the small impact on consumption. First, the project might have an impact on income sources that are correctly reported such as, for example, agricultural profit and less on income sources that are underreported, like, for example, petty trading. In this case the impact of MV on household income might appear larger than it actually is. Second, households might be saving rather than spending their income gains. For example, households may increase purchases of durable goods, productive assets or animal stocks, which we do not include in the expenditure figures.

Table 21. Difference-in-difference impact on per capita income (IPW method)

	Cross-section	Fixed effects	Lagged model
Average DD effect	0.398***	0.464***	0.210**
	(0.000)	(0.000)	(0.013)
DD effect second year	0.391***	0.414***	0.209*
	(0.000)	(0.000)	(0.054)
DD effect third year	0.405***	0.423***	0.212**
	(0.000)	(0.000)	(0.042)
Sample size	5,942	5,941	3,941

P-values in parenthesis based on cluster adjusted standard errors. *** is statistical significance at 1%, ** is 5% significance and * is 10% significance.

We further investigate the impact of the intervention on household income by disaggregating changes by income source. Changes in agricultural profits would suggest that the project is effective in increasing agricultural productivity through training and modern inputs. Changes in livestock income would be more difficult to interpret as the project does not contemplate initiatives directed to increase income from livestock. An increase in business income in MV would suggest an increase in general economic activity possibly stimulated by the injection of liquidity and facilitated by the easier access to markets through improvement in infrastructure. A change in income from employment would suggest that the project is increasing incomes of those directly employed by the intervention or indirectly through the stimulus to economic activity. Finally, an increase in transfers would be hard to interpret as the project is not encouraging migration and does not provide direct monetary transfers. To simplify the presentation of the results we only show the more conservative estimates obtained using the lagged dependent variable model (Table 22).

Table 22. Impact of MV on different income sources (lagged model)

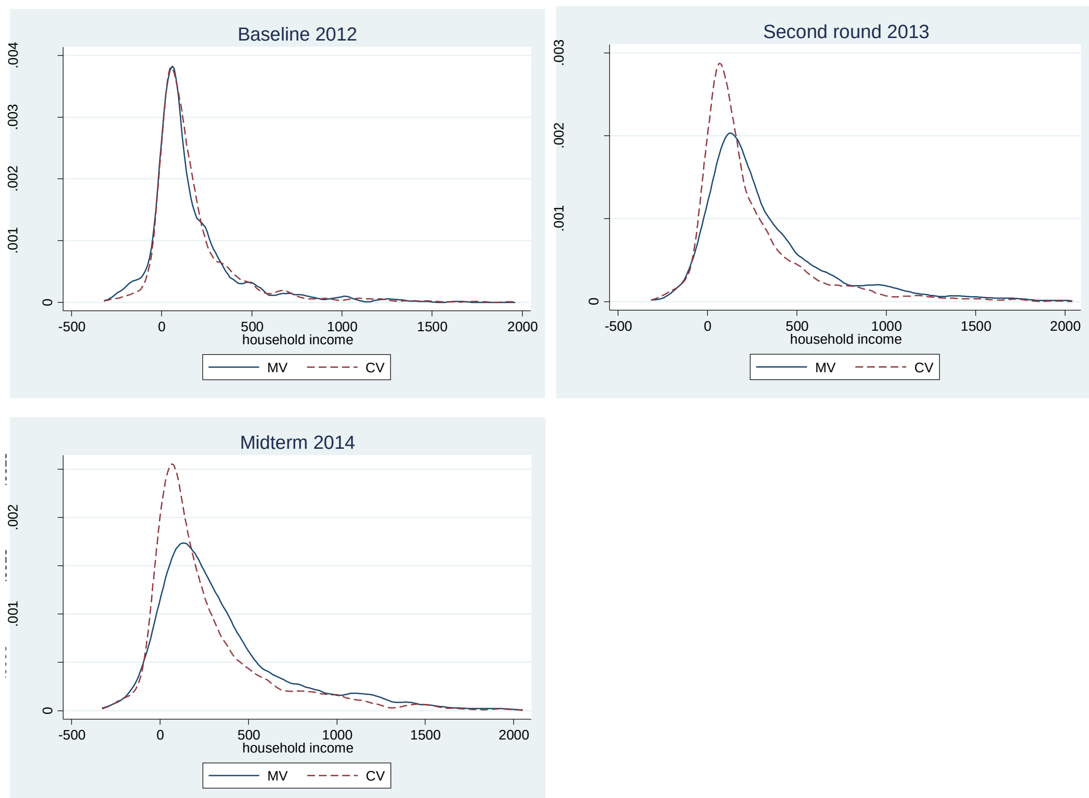
	Average DD effect	DD effect second year	DD effect third year
Agricultural income	0.264**	0.261**	0.266
	(0.041)	(0.038)	(0.101)
Livestock income	0.207**	0.044	0.241**
	(0.015)	(0.687)	(0.006)
Business income	0.244**	0.327**	0.161
	(0.016)	(0.046)	(0.163)
Employment income	-0.087*	-0.046	-0.127
	(0.078)	(0.196)	(0.141)
Transfers income	0.114***	0.066	0.163**
	(0.000)	(0.245)	(0.021)

P-values in parenthesis based on cluster adjusted standard errors. *** is statistical significance at 1%, ** is 5% significance and * is 10% significance.

The MV appears to change all income sources. The largest increases are observed in business income and agricultural income. Since the latter represents nearly 50% of overall income, it largely explains the overall income change observed. There is also a noticeable increase in livestock income and an increase in income from transfers. Interestingly there is a small but statistically significant reduction in income from employment as if productive resources were being transferred to agriculture from other wage activities. These results seem to suggest that the project was successful in increasing agricultural productivity and stimulating overall economic activity in the area.

We then investigate whether the intervention had a different impact at different levels of the income distribution. The charts of Figure 6 show the densities of household incomes in the project and comparison areas for the three survey rounds. The shapes of the distributions are remarkably similar in the project and control groups at the baseline. There is a clear shift to the right of the MV income distribution at the second round which becomes even more accentuated at the midterm survey. The impact of the intervention appears to be more pronounced in the middle and upper end of the income distribution.

Figure 6. Densities of household income in the project and control groups



These impressions are confirmed by quantile regression analysis which shows impacts that are higher at higher expenditure quantiles (Table 23). These results are in sharp contrast with those observed in relation to per capita expenditure where impact of the intervention appeared to be larger for the poorer sectors of the population. If the discrepancy in the project effects on income and expenditure occurs because households are saving income gains, then the discrepancy in the distribution of effects on income and expenditure could be the result of the better-off households saving gains while worse-off are spending them.

Table 23. Distributional impact on household income (average DD effects)

	10th percentile	25th percentile	50th percentile	75th percentile	90th percentile
Average DD effect	0.083**	0.157***	0.224***	0.291**	0.397**
	(0.010)	(0.000)	(0.000)	(0.002)	(0.015)
DD effect second year	0.087**	0.125***	0.185**	0.277**	0.373*
	(0.037)	(0.000)	(0.001)	(0.035)	(0.051)
DD effect third year	0.068	0.176***	0.263***	0.548**	0.430
	(0.136)	(0.000)	(0.000)	(0.027)	(0.104)

Difference-in-difference coefficients of lagged models. P-values in parentheses are corrected by clustering at the village level using bootstrap sample replications. *** is statistical significance at 1%, ** is 5% significance and * is 10% significance.

We also investigated the impact of the intervention on food security. The household surveys collect information on perceptions of food security via two questions. The first question elicits a binary response: "In the past 12 months, were there months in which you did not have enough food to meet your family's needs?" The second question obtains a continuous response: "How many days in the last 30 days did you not have enough food to meet your family's needs?" Note that these questions were asked only at the baseline and midterm rounds and not during the 'in-between' rounds so that difference-in-difference effects are reported with respect to the change between the baseline and the midterm surveys. The effects are, as usual, calculated in three difference ways and adjusted by baseline characteristics using the inverse probability method. For simplicity, we employ the same propensity score already employed when calculating the impact of the project on income and expenditure as the determinants of food security are similar to the determinants of income and expenditure.

Table 24. Project impact on food security perceptions (IPW method)

	Cross-section	Fixed effects	Lagged model
Not enough food in the last 12 months	-0.33*** (0.000)	-0.33*** (0.000)	-0.32*** (0.000)
Days without enough food in the last month	-1.9 (0.213)	-1.8 (0.211)	-2.5** (0.005)
Sample size	3,960	3,916	1,958

Difference-in-difference coefficients of lagged models. P-values in parentheses are corrected by clustering at the village level using bootstrap sample replications. *** is statistical significance at 1%, ** is 5% significance and * is 10% significance.

The project appears to produce a large percentage reduction (30%) in the fraction of the population reporting not having enough food to eat over the last 12 months (Table 24). There is also a reduction in the number of reported days during which the household did not have enough to eat over the last 30 days. The impact for this second measure of food security is somewhat smaller and not robust across different specification. Overall, these data show that perceptions of food security have improved in the MV areas.

7. Impact of the MVP on child health

In this section we examine the impact of MV on three major indicators of child health: mortality, anthropometry and anaemia.

Child mortality in Ghana

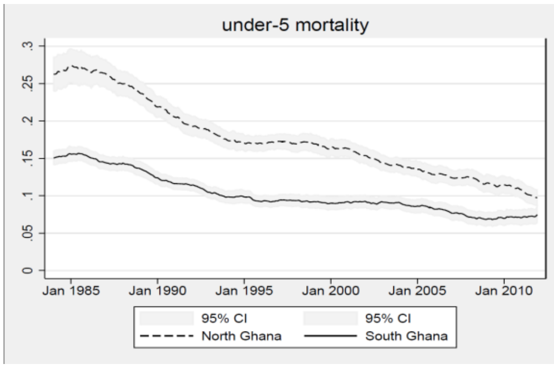
Most recent available data report the following child mortality rates for Ghana (Ghana Statistical Service 2011):

Table 25. Child mortality rates in Ghana (DHS 2011)

Mortality	Rate per thousand
Neonatal	32
Post-neonatal	21
Infant	53
Child	31
Under-5	82

Mortality rates have been decreasing in Ghana over the last 30 years. MDG4 called for a two-thirds reduction in under-5 mortality rates between 1990 and 2015. By 2011 the under-5 mortality rate was reduced by less than a third, so that the goal will not be met. The reduction in mortality rates has been much faster in the North than in the South of the country. In 1985 one in four children would die before their fifth birthday in Northern Ghana. In 2011 the same probability was one in ten, not too different from the same probability in Southern Ghana (Figure 7[3]).

Figure 7. Under-5 mortality rates in Northern and Southern Ghana 1985–2011



Baseline mortality rates in project and control areas

Baseline mortality rates of project and comparison areas are very different (Table 26) (details of the calculations and quality checks of the data are in Appendix 4). The sample sizes are relatively small and we calculated conservative standard errors assuming dependence of observations within clusters. As a result, only two differences (post-neonatal and under-5 mortality) are significantly different at 10%.

[3] Figure 7 was obtained by pooling the data from six different DHS surveys: 1988, 1993, 1998, 2003, 2008 and 2011. Mortality rates were calculated for each of 300 months before the survey in 2011 exploiting mothers' retrospective recall of births and deaths. When pooling data from different samples a problem arises about the use of the existing sampling weights. While several and complex procedures are possible (Korn and Graubard), we use here the simplest reweighting scheme consisting of adjusting the sampling weight in each survey by the sample size contribution of the survey to the sample of pooled surveys. For example, the sampling weights of the 1988 survey with sample size N_{88} are obtained by multiplying the sampling weights of the 1988 dataset by the ratio N_{88}/N, where N is the sample size of all pooled datasets. The large difference in under-5 mortality in 1985 narrowed down dramatically over the period considered.

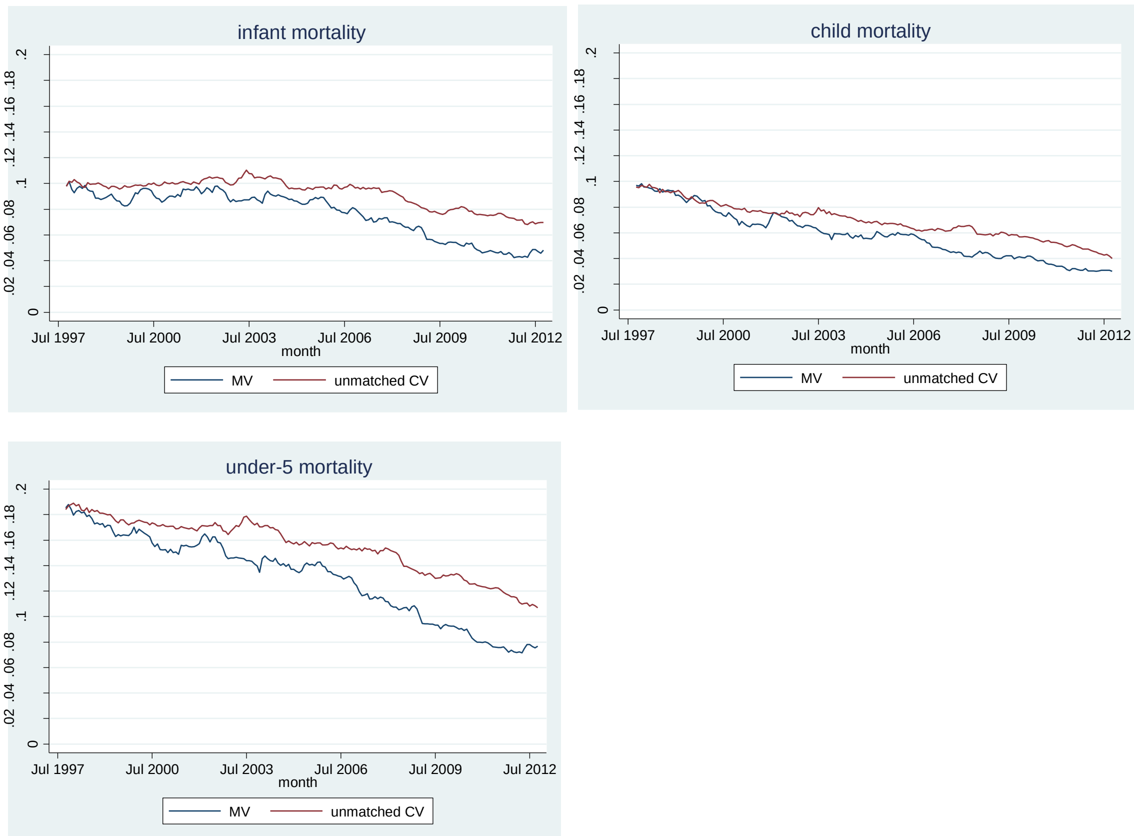
Table 26. Differences in baseline mortality rates in MV and CV areas

Mortality rate	Project	Control	Difference
Neonatal	**29.43**	**39.12**	**-9.69**
95% CI	(11.06, 49.31)	(27.61, 50.00)	(-13.90, 33.13)
S.e.	9.94	5.85	11.61
P-value	0.003	0.000	0.404
Post-neonatal	**18.67**	**30.53**	**-11.86***
95% CI	(10.82, 27.96)	(20.71, 40.82)	(-1.39, 24.72)
S.e.	4.32	5.01	6.73
P-value	0.000	0.000	0.078
Infant	**48.10**	**69.65**	**-21.54**
95% CI	(26.09, 72.14)	(52.31, 87.18)	(-1.15, 52.44)
S.e.	11.85	9.06	15.27
P-value	0.000	0.000	0.158
Child	**30.15**	**40.11**	**-9.96**
95% CI	(18.19, 39.33)	(30.34, 51.23)	(-4.82, 26.16)
S.e.	5.35	5.33	7.62
P-value	0.000	0.000	0.191
Under-5	**76.80**	**106.96**	**-30.16***
95% CI	(54.32, 101.48)	(84.91, 129.65)	(-3.86, 62.97)
S.e.	12.23	11.37	17.03
P-value	0.000	0.000	0.077

Mortality rates are calculated using the synthetic cohort probability method used by the DHS. Standard errors are calculated using bootstrap replications. *** is statistical significance at 1%, ** is 5% significance and * is 10% significance.

What is more striking of the differences between MV and CV areas is that while CV mortality rates are similar to those prevailing in rural areas of the Northern Ghana, mortality rates in MV areas are much lower. We conducted field interviews to understand the reason for this differences and we concluded that it could be a consequence of activities carried out in the area by the Ghanaian Health Service in collaboration with the Navrongo Health Research Center established in Kasena-Nangana (bordering North to the Builsa district) in the early 1990s with the support of the London School of Hygiene and Tropical Medicine. Other NGOs have operated in the area and may have contributed to improving child mortality such as the Wiaga clinic, run by the Christian Health Association of Ghana, and the community Radio Builsa promoting behavioural change in health and education from the town of Sandema. Whatever the reason for the low mortality rates in the project areas, the CV areas do not provide a good comparator group for the analysis. This is further illustrated by an analysis of mortality trends in the two areas. The charts in Figure 8 show the trend in infant, child and under-5 mortality rates over the 15 years before the baseline survey. Mortality rates were similar in the two areas until 10 years before the survey but started to diverge in the mid-2000s. The difference in the levels and, possibly, in the trends is more apparent in the case of under-5 mortality.

Figure 8. Mortality trends in MV and CV compared (1997-2012)



Matching of project and control villages at the design stage built a comparison group of villages similar to those in the project group with respect to most socio-economic outcome indicators. However, village-level matching failed to build a valid control group with respect to child mortality rates. Therefore, unlike in the analysis of income and expenditure conducted in the previous sections, we want to be able to remove differences in characteristics at the household and village level. The goal is to achieve statistical balance in mortality outcomes between the project and the comparator group. To do so, we run a logit participation model, which includes household-level and village-level determinants of selection in the project and of mortality. The results are reported in Table 27.

Table 27. Probit selection model for mortality observations

Variable	Coeff.	Standard Error	P-value
Mother's age	-0.015	0.003	0.000
Number of births	0.004	0.010	0.716
Literacy score	-0.079	0.016	0.000
Married	-0.257	0.122	0.035
Domestic violence	-0.147	0.044	0.001
Sexual violence	-0.136	0.050	0.006
Household size	-0.035	0.004	0.000
Cement floor	-0.096	0.044	0.028
Metal roof	0.680	0.039	0.000
No windows in the home	-0.182	0.042	0.000
No place for hand washing	0.144	0.053	0.007
Improve sanitation	-0.003	0.058	0.954
Flood in previous 3 years	0.136	0.015	0.000
Number of bedrooms	0.048	0.007	0.000
Asset index	1.227	0.166	0.000
Village average of Builsa group	0.375	0.177	0.034
Village average of Mampruli group	0.254	0.194	0.191
Village distance to health clinic	-0.069	0.002	0.000
Women group in the village	0.855	0.043	0.000
Health committee in the village	0.360	0.041	0.000
Constant	-1.282	0.270	0.000
Pseudo R-square			0.273
Observations			9,337

Village-level variables are highly statistically significant. We use the propensity score to weight observations inversely to selection probability and we check the statistical balance of mortality outcomes. After matching, the differences in mortality rates between the MV observations and the matched CV observations are minimal and the P-values of significance test of the differences are close to one in most cases.

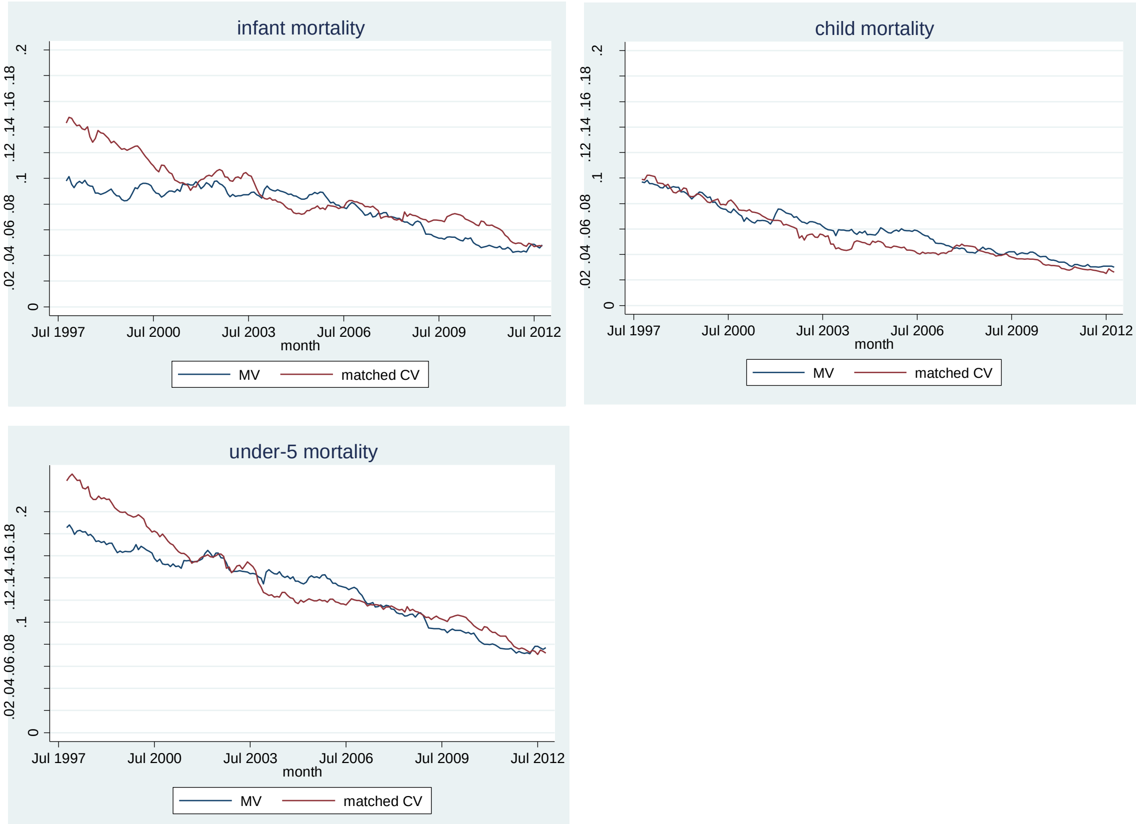
Table 28. Adjusted differences in baseline mortality rates in MV and CV areas

Mortality rate	Project	Control	Difference
Neonatal	**29.43**	**29.07**	**0.36**
95% CI	(11.06, 49.31)	(13.04, 58.27)	(-26.52, 32.49)
S.e.	9.94	11.88	14.78
P-value	0.003	0.000	0.980
Post neonatal	**18.67**	**18.27**	**0.40**
95% CI	(10.82, 27.96)	(7.19, 37.97)	(-14.79, 17.27)
S.e.	4.32	7.96	8.41
P-value	0.000	0.000	0.962
Infant	**48.10**	**47.34**	**0.76**
95% CI	(26.09, 72.14)	(25.21, 85.26)	(-34.85, 37.27)
S.e.	11.85	15.42	18.14
P-value	0.000	0.000	0.967
Child	**30.15**	**26.01**	**4.14**
95% CI	(18.19, 39.33)	(15.36, 41.73)	(-19.03, 14.17)
S.e.	5.35	6.78	8.49
P-value	0.000	0.000	0.626
Under-5	**76.80**	**72.12**	**4.68**
95% CI	(54.32, 101.48)	(43.24, 118.88)	(-42.83, 41.88)
S.e.	12.23	19.81	21.74
P-value	0.000	0.000	0.829

Mortality rates are calculated using the synthetic cohort probability method used by the DHS. Standard errors are calculated using bootstrap replications. Observations are weighted using inverse probability weights derived from the propensity score based on the model of Table 26. *** is statistical significance at 1%, ** is 5% significance and * is 10% significance.

The similarity between the MV and matched control observations is most visible when comparing the trends in mortality rates over the 15 years before the baseline. Figure 9 shows that levels and trends of the mortality series in project and comparison areas using the adjusted data are nearly identical over the 10 years before the baseline.

Figure 9. Mortality trends in MV and matched CV compared



Next we calculate the difference-in-differences in mortality rates between the baseline and the midterm using the adjusted data. The results are reported in Table 29 and have the expected sign. Mortality rates are decreasing in MV areas at a higher speed than in CV areas. The difference however is small and it is not statistically significant. This is partly the result of the small sample size but, more importantly, of the fact that the mortality rates reported in Table 29 are calculated over five years before the survey. Rates calculated in this way are not appropriate for a difference-in-difference analysis because the rates calculated at midterm (two years after project start) also include birth and death events occurred up to three years before the baseline. Only a dramatic change in mortality rates after the baseline would be visible using these data. The data suggest that a drop in mortality rates occurred but better estimates will be calculated in 2016 when the fifth round of data collection will be available and mortality rates of four or five-year intervals could be calculated and compared over time between MV and CV areas.

Table 29. Difference-in-difference analysis (five-year interval)

Mortality rate	Baseline difference	Midterm difference	Difference-in-difference
Neonatal	**0.36**	**0.89**	**0.53**
95% CI	(-26.52, 32.49)	(-10.71, 11.51)	(-23.55, 32.77)
S.e.	14.78	5.73	13.87
P-value	0.980	0.877	0.970
Post-neonatal	**0.40**	**-6.00**	**-6.40**
95% CI	(-14.79, 17.27)	(-14.99, 35.90)	(-34.15, 14.34)
S.e.	8.41	12.63	12.19
P-value	0.962	0.635	0.600
Infant	**0.76**	**-5.11**	**-5.87**
95% CI	(-34.85, 37.27)	(-15.63, 35.77)	(-43.18, 32.32)
S.e.	18.14	13.07	19.73
P-value	0.967	0.696	0.766
Child	**4.14**	**-6.47**	**-10.62**
95% CI	(-19.03, 14.17)	(-10.21, 23.60)	(-30.78, 14.40)
S.e.	8.49	8.71	11.40
P-value	0.626	0.457	0.354
Under-5	**4.68**	**-11.26**	**-15.95**
95% CI	(-42.83, 41.88)	(-16.81, 47.96)	(-60.07, 33.83)
S.e.	21.74	16.39	24.35
P-value	0.829	0.492	0.512

Mortality rates are calculated using the synthetic cohort probability method used by the DHS. Standard errors are calculated using bootstrap replications. Observations are weighted using inverse probability weights derived from the propensity score based on the model of Table 26. *** is statistical significance at 1%, ** is 5% significance and * is 10% significance.

Impact of MV on anthropometry

The project had a positive impact on physical growth of children. Anthropometric data were only collected at the baseline and at the midterm so that the difference-in-difference analysis can only be conducted between two points in time. We look at standardised Z-scores among children under-5: height-for-age, weight-for-age and weight-for-height (Table 30). Scores were calculated using the World Health Organisation (WHO) stata codes employing most recent reference growth charts. The difference-in-difference coefficients are adjusted for differences in baseline characteristics using inverse probability weights. Note that we corrected for baseline characteristics of households rather than baseline characteristics of children because many of the children measured at the midterm had not been measured at the baseline. For the same reason, we do not conduct a panel fixed effects analysis because it could only be conducted on a fraction of the available sample.

Table 30. Impact of the MVP on nutritional status of children under-5 (IPW method)

	Difference-in-difference	P-value	Sample size
Height-for-age	0.48***	0.001	3,392
Weight-for-age	0.24**	0.036	3,398
Weight-for-height	-0.11	0.373	3,384

*** is statistical significance at 1%, ** is 5% significance and * is 10% significance.

The observed impacts of the project on weight-for-age and particularly on height-for-age are very large for a typical intervention focused on agriculture (Masset *et al.* 2012) and it is large even for specific nutrition interventions (IEG 2010). Heights and weights of children have improved in MV areas and since they have got slightly worse in comparison areas over the same period, the difference-in-difference effect is magnified. Nutrition is determined by multiple factors including: the size and the composition of the diet, morbidity and general health environment, and parental care. All these factors may have contributed to a varying extent to achieve this result. A more detailed analysis, unpacking the determinants of these results, is in order but beyond the scope of this report.

Impact of the MVP on anaemia

The baseline survey found very high prevalence rates of anaemia that are comparable to those prevailing in Ghana and other West African countries. Following DHS standards, we calculate mild anaemia prevalence as the ratio of children with haemoglobin below 11 g/dL. Moderate anaemia is haemoglobin below 10 g/dL and severe anaemia is haemoglobin below 7 g/dL. We find no improvement on average haemoglobin concentration or in the distribution of anaemia prevalence (Table 31). Haemoglobin concentration is slightly higher in MV areas at midterm and the prevalence of anaemia is slightly smaller in MV areas, but these differences are not statistically significant.

It should be noted however that these results could have been biased by seasonal differences. Baseline blood tests were conducted during the dry season in the MV areas (May–June), when haemoglobin concentration is higher, and in the wet season in the CV areas when haemoglobin concentration is lower (August–September). The midterm data were collected in the period July–August in both MV and CV areas. It is possible that the improvement over time observed in the CV areas is partly the result of seasonal bias and that the improvement observed in the MV areas would have been much larger had the blood test being conducted in the wet season rather than in the dry season.

Table 31. Impact of the MVP on haemoglobin concentration and prevalence of anaemia

	Baseline			Midterm				
Indicator	MV	CV	Diff	MV	CV	Diff	DD	DD*
Haemoglobin	9.98	9.46	0.51**	10.2	9.98	0.20*	-0.31	-0.32
P-value			(0.001)			(0.054)	(0.149)	(0.171)
Mild anaemia	74.0	84.3	-10.3**	72.1	74.3	-2.1	8.1	8.2
P-value			(0.003)			(0.498)	(0.128)	(0.182)
Moderate anaemia	45.7	61.9	-16.2***	43.2	47.0	-3.7	12.4*	12.1
P-value			(0.000)			(0.324)	(0.064)	(0.102)
Severe anaemia	3.9	5.2	-1.34	1.0	2.0	-1.0	0.3	1.1
P-value			(0.464)			(0.200)	(0.859)	(0.594)

Unadjusted difference-in-differences (DD) and adjusted (DD*) are reported in the last two columns *** is statistical significance at 1%, ** is 5% significance and * is 10% significance.

8. Impact of the MVP on education

The project appears to have increased school attendance (Table 32). However, this has not resulted in an improvement in learning and cognitive abilities as measured by cognitive tests and Math and English test scores. We first conduct a difference-in-difference analysis of the impact of MV on school attendance of primary, junior secondary and senior secondary school. The analysis is cross-sectional because only a fraction of the sample children overlap across the two surveys and restricting the analysis to the sample of panel children would heavily reduce the sample size. For the same reason we adjust for differences in baseline characteristics at the household level rather than at the child level. The largest project impact occurred in primary school (8% increase in attendance). A smaller impact (5% increase) is observed in junior secondary school and no impact is found in senior secondary. The impact appears to have been larger in the first year of the project in primary school and in the second year of the intervention in junior secondary.

Table 32. Impact of the MVP on school attendance

	Primary	Junior Secondary	Senior Secondary
Average DD effect	0.08**	0.05*	0.00
	(0.029)	(0.099)	(0.799)
DD effect second year	0.11**	0.03	0.001
	(0.006)	(0.371)	(0.735)
DD effect third year	0.047	0.08*	0.00
	(0.205)	(0.052)	(0.909)
Sample size	8,022	3,161	4,019

*** is statistical significance at 1%, ** is 5% significance and * is 10% significance.

Cognitive skills did not improve over the two-year period of intervention (Table 33). While the test scores on Raven's matrices have improved in MV areas, the forward and backward digit span test show a regress in MV areas compared to CV areas. No impact is visible on simple English and Math tests, though there is a sizable improvement in advanced English and Math tests, which are administered to children attending junior secondary school. Relatively poor test score results cannot be explained by the fact that MV is bringing into school children from disadvantaged backgrounds that are on average performing worse than children attending school in the control group. Cognitive tests, as well as simple Math and English tests, were conducted at the household level regardless of schooling level so that this type of bias is unlikely.

Table 33. Impact of the MVP on test scores

	Difference-in-difference	P-value	Sample size
Raven's matrices	0.18	0.164	6,565
Forward digit span	-0.19*	0.076	6,471
Backward digit span	-0.22*	0.097	6,467
Easy English	-0.03	0.771	3,204
Easy Math	0.03	0.729	3,571
Advanced English	0.44**	0.004	0.929
Advanced Math	0.27	0.105	0.920

*** is statistical significance at 1%, ** is 5% significance and * is 10% significance.

9. Impact heterogeneity

In this section we analyse the impact of the intervention by district and gender. The population reached by the project is equally distributed between the original districts of Builsa and West Mamprusi. After the creation of new districts in 2012, the MV project villages are distributed in the following way: 23 in Builsa South, seven in West Mamprusi and five in Momprugu Moagduri. The split of the Builsa district into Builsa North and Builsa South has not particularly affected our study design as all the project villages are located in Builsa South. The split of the West Mamprusi district on the other hand has divided project communities almost equally between the two newly created districts, which are administered in different ways. In our analysis, however, we employ the original subdivision pre-2012 between Builsa and West Mamprusi districts. This subdivision has only a partial validity in terms of administrative and political differences between the two areas but it does characterise areas that are quite homogeneous in terms of language spoken, ethnic groups, socio-economic characteristics and social and political organisation.

Table 34. Impact of the MVP by district

	Builsa	West Mamprusi	F-test
Per-adult equivalent expenditure	0.097 (0.202)	-0.020 (0.799)	1.65 (0.202)
Per-capita income	0.518*** (0.000)	0.313*** (0.000)	2.91* (0.091)
Food security (1)	-0.307*** (0.000)	-0.346*** (0.000)	0.22 (0.642)
Food security (2)	-2.411 ((0.197)	-1.225 (0.369)	0.62 (0.431)
Height-for-age	0.673*** (0.000)	0.374*** (0.000)	2.27 (0.135)
Weight-for-age	0.292** (0.025)	0.218 (0.133)	0.22 (0.641)
Weight-for-height	-0.135 (0.150)	-0.088 (0.608)	0.08 (0.778)
Haemoglobin concentration	-0.168 (0.443)	-0.436 (0.199)	0.59 (0.443)
Mild anaemia	0.052 (0.439)	0.104 (0.218)	0.31 (0.579)
Moderate anaemia	0.087 (0.286)	0.147 (0.143)	0.31 (0.579)
Severe anaemia	-0.001 (0.951)	0.021 (0.141)	0.78 (0.378)
Attendance primary school	0.030 (0.655)	0.106** (0.003)	1.12 (0.293)
Attendance junior secondary school	0.013 (0.746)	0.083* (0.052)	1.77 (0.187)
Attendance senior secondary school	0.015 (0.417)	-0.003 (0.881)	0.67 (0.415)
Raven's test	0.224* (0.052)	0.188 (0.364)	0.03 (0.857)
Digit forward test	-0.135 (0.322)	-0.169 (0.189)	0.04 (0.838)
Digit backward test	-0.361** (0.005)	-0.185 (0.311)	0.86 (0.357)
Easy math test	0.089 (0.453)	0.034 (0.745)	0.22 (0.642)
Easy English	-0.019 (0.871)	0.174 (0.289)	1.19 (0.278)
Advanced math	-0.030 (0.881)	0.718*** (0.000)	13.06** (0.001)
Advanced English	0.355* (0.094)	0.769*** (0.000)	3.23* (0.076)

The table reports difference-in-difference coefficients of regression analysis based on cross-sectional models adjusted by inverse probability weights. *** is statistical significance at 1%, ** is 5% significance and * is 10% significance.

In order to assess whether the project had a different impact in the two districts we consider the Builsa and West Mamprusi MV villages as two different interventions. As usual we calculate difference-in-difference effects using regression analysis after weighting observations for their selection probability. Table 34 shows the difference-in-difference effects estimated for the two districts in relation to the outcomes analysed in the previous sections, together with an F statistics that tests the difference between the coefficients. The project appears to have a larger impact in Builsa than in West Mamprusi for most economic and health indicators, though a statistical test of the difference between the coefficients is significant only in the case of per-capita income. Interestingly, the project impact is stronger in West Mamprusi than in Builsa for education indicators (attendance rates and test scores), though again the difference between the coefficients is never statistically significant. Overall the data of Table 34 suggest that the project was moderately more effective in Builsa in promoting child health and economic activities, while it was more effective in West Mamprusi in promoting child education.

We conduct the same type of analysis disaggregating the impact of the intervention by gender for all individual-level outcomes considered in the previous sections (Table 35). We find no differences in project impact on nutritional status between boys and girls. Interestingly, the programme appears to be more effective in improving anaemia among girls than boys. On the other hand, the programme appears to be more effective in improving primary school attendance of boys. Overall the data suggest that project impact is quite evenly distributed between boys and girls and that differences emerge in those cases where boys and girls have different initial status. For example, more girls are normally attending school than boys, which might explain the project impact on school attendance of boys.

Table 35. Impact of the MVP by gender

	Boys	Girls	F-test
Height-for-age	0.483** (0.003)	0.510** (0.001)	0.04 (0.849)
Weight-for-age	0.264* (0.053)	0.233* (0.060)	0.07 (0.794)
Weight-for-height	-0.084 (0.539)	-0.128 (0.297)	0.21 (0.648)
Haemoglobin concentration	-0.448* (0.086)	-0.158 (0.491)	3.97* (0.049)
Mild anaemia	0.117 (0.100)	0.038 (0.559)	1.82 (0.180)
Moderate anaemia	0.155* (0.054)	0.079 (0.307)	1.95 (0.166)
Severe anaemia	0.017 (0.515)	0.002 (0.936)	0.52 (0.471)
Attendance primary school	0.097** (0.020)	0.038 (0.313)	2.55 (0.113)
Attendance Junior secondary school	0.057* (0.080)	0.050 (0.277)	0.03 (0.864)
Attendance senior secondary school	0.010 (0.562)	-0.07 (0.813)	0.37 (0.547)
Raven's test	0.257 (0.111)	0.149 (0.309)	0.98 (0.324)
Digit forward test	-0.212* (0.053)	-0.098 (0.381)	2.63 (0.108)
Digit backward test	-0.287* (0.033)	-0.221 (0.122)	0.69 (0.409)
Easy math test	-0.015 (0.886)	0.148 (0.183)	1.76 (0.187)
Easy English	0.002 (0.991)	0.211** (0.047)	1.70 (0.195)
Advanced math	0.265 (0.127)	0.341* (0.084)	0.22 (0.642)
Advanced English	0.586** (0.005)	0.518** (0.015)	0.06 (0.803)

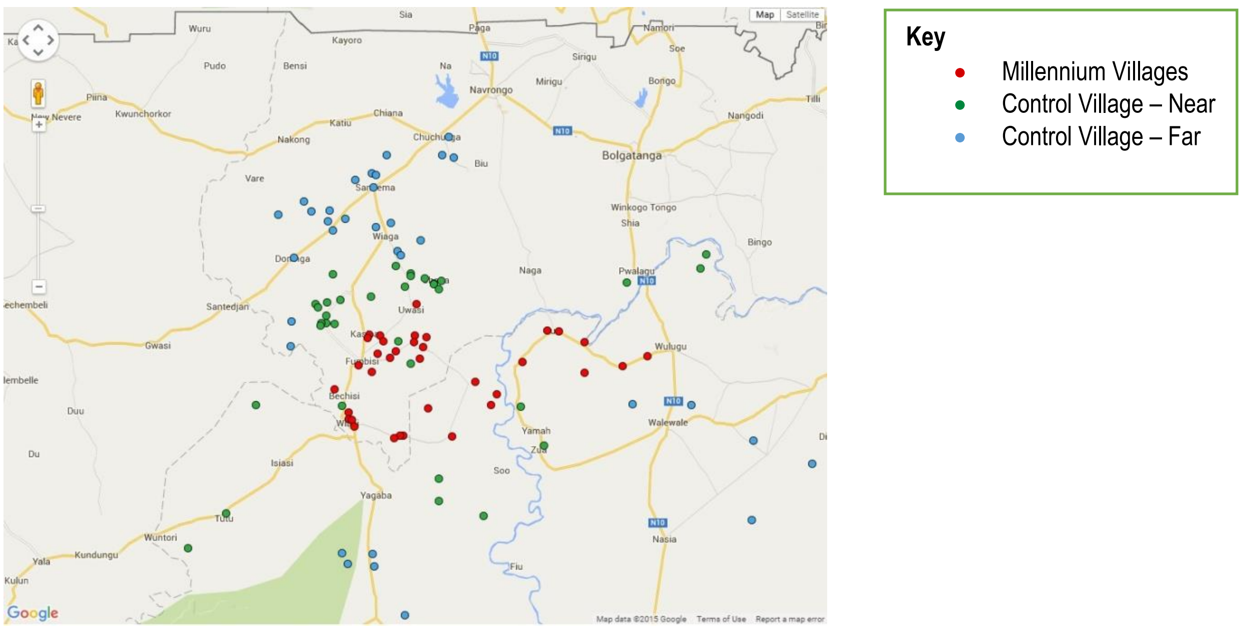
The table reports difference-in-difference coefficients of regression analysis based on cross-sectional models adjusted by inverse probability weights. *** is statistical significance at 1%, ** is 5% significance and * is 10% significance.

10. Spill-over effects

In this section we investigate the presence of spill over effects that is whether the MV project had an impact beyond the targeted area of intervention. The detection of spill over effects would imply that the impacts estimated in the previous sections are underestimates of true project effects. As discussed in the analysis plan, this analysis is exploratory as we have not formulated a conceptual framework describing the mechanisms through which spill over effects should operate. There have been reports that families residing in non-project areas are covering long distance to access health clinics run by the project, and several channels could be hypothesised through which the programme could have an economic impact on areas outside the area of intervention through, for example, markets and prices.

We use the distance of control group villages from MVs as proximate determinant of spill over effects but we do not have any guiding principle to establish whether any given distance should be considered ‘near’ or ‘far’. The original sample was stratified by distance, whereby villages were selected in equal proportions (within each district) from villages within an area at 15 to 20 kilometre distance from the Millennium Villages and from villages beyond this distance. This stratification is depicted in Figure 10, where the red dots are MV villages, the green dots are ‘nearby’ comparison villages and the blue dots are ‘far-away’ comparison villages. This stratification was not very accurate because it was not based on GPS coordinates, which were not available at the time. Hence, in addition to the sample design original subdivision between ‘near’ and ‘far’ villages we consider a second subdivision which employs GPS coordinates to calculate distances and that weight distances by population density in the MV villages. This second subdivision consists of a weighted average of the distance of each comparison village from all other MVs weighted by their population. The idea behind the weighting down of poorly populated MV villages is that any spill over effect should be larger from a larger project community than from a smaller one.

Figure 10. Project and comparison communities[4]



To calculate spill over effects we split the sample of comparison villages into two equal groups based on the two criteria of ‘near’ and ‘far’ outlined above. We then estimate the project effects in the usual way: using difference-in-difference analysis and correcting for differences in baseline characteristics by inverse probability weights. We focus first on outcomes for which we know, from the previous analysis, that the project had an impact: per capita income and anthropometry. Table 36 shows difference-in-difference coefficients for MV areas and ‘nearby’ areas using the four definitions described above. Impacts in MV areas and ‘near’ comparison villages are estimated using ‘far’ comparison villages as the comparison group. The impact of MV on income is larger once neighbouring villages are considered as part of the intervention area. The impact in neighbouring villages (spill over effects) appears to be about half the size of project effects. The impact is transmitted through all income sources except income from transfers.

[4] Source: http://www.copypastemap.com/ to plot coordinates from the Earth Institute on Google Maps.

The two difference distance measures employed seem to capture the same phenomenon and the differences between the coefficients are rather small.

Table 36. Impact of the MVP on income by distance

	Original 'near' and 'far' subdivision		Population weighted average distance subdivision	
	MV	Near	MV	Near
Income	0.48***	0.30**	0.49***	0.27**
	(0.000)	(0.001)	(0.000)	(0.004)
Agricultural income	0.045***	0.31	0.46***	0.23
	(0.000)	(0.103)	(0.000)	(0.257)
Livestock income	0.030**	0.14**	0.30**	0.15**
	(0.001)	(0.009)	(0.001)	(0.004)
Business income	0.53***	0.27***	0.52***	0.23***
	(0.000)	(0.000)	(0.000)	(0.000)
Employment income	-0.04	0.04	-0.04	0.05
	(0.401)	(0.442)	(0.422)	(0.411)
Transfers income	0.11**	0.04	0.10**	-0.01
	(0.029)	(0.917)	(0.032)	(0.901)

The table reports difference-in-difference coefficients of regression analysis based on cross-sectional models adjusted by inverse probability weights. *** is statistical significance at 1%, ** is 5% significance and * is 10% significance.

We find no spill over effects of the intervention on nutritional status of children. Table 37 shows difference-in-difference estimates of project impact on anthropometric indicators by distance. In this case we add another distance classification based on the distance of the comparison community from the nearest clinic supported by MV (last column of Table 37). The project does not have an impact on nutritional status of children in areas in the vicinity of MV clinics or in the geographical vicinity of MV villages as defined by the classification described above.

Table 37. Impact of the MVP on nutritional status by distance

	Original 'near' and 'far' subdivision		Population weighted average distance subdivision		Subdivision based on distance from nearest clinic	
	MV	Near	MV	Near	MV	Near
Height-for-age	0.50**	0.01	0.49**	-0.08	0.49**	0.04
	(0.004)	(0.983	(0.012)	(0.679)	(0.007)	(0.983
Weight-for-age	0.12	-0.25*	0.07	-0.036**	0.11	-0.26*
	(0.312)	(0.097)	(0.618)	(0.012)	(0.456)	(0.098)
Weight-for-height	-0.30*	-0.38**	-0.30**	-0.41***	-0.26*	-0.30**
	(0.044)	(0.006)	(0.027)	(0.000)	(0.063)	(0.025)

Appendix A1. Definitions of MDG indicators

Proportion of population below $1 (PPP) per day. The proportion of the population living in households below the international *poverty line* where the average daily consumption per person is less than PPP $1.25 a day. PPP dollars are calculated using the PPP conversion factors reported by the World Bank in the world development. The series for Ghana are: 2011, 0.699; 2012, 0.792; 2013, 0.916; 2014, 1.035. The purchasing power parity conversion factor is the number of units of a country's currency required to buy the same amounts of goods and services in the domestic market as US dollar would buy in the United States. The PPP US dollars equivalent of local currency per person per day was obtained by dividing per capita household consumption in Cedis by the conversion factor. In the calculations we take the average of the years 2011 and 2012 for the baseline (0.745) and the average of the years 2013 and 2014 for the midterm (0.976)

Proportion of population below the national poverty line. The proportion of the total population living below the *national poverty line.* We followed official Ghanaian statistics in defining the poverty line and calculating the poverty headcount. A person is classified as poor if living in a family whose expenditure per-adult equivalent is below 1,313 Cedis per year at 2013 Accra prices (equivalent to US $1.84 per person per day and PPP US $3.93 per person per day – using average exchange rates and PPP conversion factors of 2013 of 1.95 Cedis per dollar and 0.916, respectively). Following GSS, household expenditures were adjusted for regional differences in prices and for inflation using the adjustment factors available with GLSS6.

Poverty gap ratio. The mean shortfall of the total population from the *poverty line* (counting the non-poor as having zero shortfall), expressed as a percentage of the *poverty line* as defined above.

Share of poorest quintile in national consumption. This is the share of a country's national consumption or income that accrues to the *poorest quintile* (fifth) of the population. In our application we use the share of total expenditure in the study area (rather than the national expenditure) that accrues to the poorest quintile of the population ranked by per-adult equivalent expenditure

Employment to population ratio. This is the proportion of a country's working-age population (15 and older) that is employed. We use instead the percentage of individuals in the study population older than 15 who did any work, paid or unpaid, over the previous year. This definition does not include domestic work.

Proportion of employed people living below $1 (PPP) per day. This is the proportion of individuals who are *employed* but nonetheless live in a household whose members live below the *poverty line.* In our application we use the percentage of the employed (any individual above the age of 15 who did any work, paid or unpaid, over the previous year) whose households have a per-capita expenditure below PPP US $1.25. The employed and the PPP US$ were defined above.

Proportion of own account and contributing family workers in total employment. This is the proportion of workers in *self-employment* who do not have *employees* and *unpaid family* workers in total *employment*. In our application we use the proportion of individuals above age 15 who did any work, paid or unpaid, within or outside the family, over the previous year either in farming, animal husbandry, fishery or any other self-employment without being remunerated.

Percentage of underweight children under-5. Children aged 0–59 months, whose weights are less than two standard deviations below the median weight for age groups in the *international reference population*. Z-scores were calculated using the *igrowup* stata programme of the WHO using the 2003 WHO growth charts.

Proportion of population below minimum level of dietary energy consumption. This is the proportion of the population whose food consumption is below a minimum dietary energy requirement for maintaining an acceptable minimum body size, a healthy life and carrying out light physical activity. In our application we use the proportion of individuals below the food poverty line (a minimum basket of food items) defined by official national statistics. Following official Ghanaian statistics a person is classified as food-poor if living in a family whose expenditure per-adult equivalent is below 792 Cedis per year at 2013 Accra prices (equivalent to US $1.11 per person per day and PPP US $2.37 per person per day – using average exchange rates and PPP conversion factors of 2013 of 1.95 Cedis per dollar and 0.916, respectively). Following GSS, household expenditures were adjusted for regional differences in prices and for inflation using the adjustment factors available with GLSS6.

Net enrolment ratio in primary education. This is the number of *children of official primary school age* who are enrolled in *primary education* to the total population of *children of official primary school age.* In our application it is the proportion of children aged 6–11 that are reported having attended primary school at any time during the previous year at the time of the interview.

Proportion of pupils starting grade 1 who reach last grade of primary. In our application this is the proportion of children aged 11–14 who ever attended primary and that completed primary.

Literacy rate of 15–24 year olds, women and men. This is defined as the proportion of the population aged 15–24 who can both read and write with understanding a short simple statement on everyday life. In our application we considered the proportion of 15–24 year olds (women and men, though mostly women as this data is part of the adult questionnaire particularly targeting women) who were able to read correctly two simple English questions (‘The child is playing with the ball’ and ‘Farming is hard work’) and were able to answer two simple arithmetic calculations (9+4 and 4x5).

Ratio of girls to boys in primary education. The *ratio* of girls to boys in *primary education, or Gender Parity Index*, is the ratio between the Gross Enrolment Ratio (GER) of girls and that of boys. In our application we used the net attendance rate in primary school of boys and girls aged 6–11 as defined above.

Ratio of girls to boys in secondary education. The *ratio* of girls to boys in *secondary education, or Gender Parity Index*, is the ratio between the GER of girls and that of boys. In our application we used the net attendance rate in junior high school of boys and girls aged 12–14 as defined above.

Ratio of girls to boys in tertiary education. The *ratio* of girls to boys in *tertiary education, or Gender Parity Index*, is the ratio between the GER of girls and that of boys. In our application we used the net attendance rate in senior secondary school of boys and girls aged 15–18 as defined above.

Share of women in wage employment in the non-agricultural sector. This is the percentage of female workers in total *wage employment* in the *non-agricultural sector*. In our application we use the proportion of women above 15 years of age who do paid work outside the agricultural sector out of the total number of individuals in the same sector.

Under-5 mortality rate. It is the child probability of dying before age five calculated over the five years preceding the interview. Rates are expressed in percentages and calculated using the synthetic cohort probability methods of employed by Demographic and Health Survey (DHS) using the *syncmrates* stata package developed by the author.

Infant mortality rate. It is the child probability of dying before age 12 months calculated over the five years preceding the interview. Rates are expressed in percentages and calculated using the synthetic cohort probability methods of employed by DHS using the syncmrates stata package developed by the author.

Proportion of 1-year-old children immunised against measles. It is the proportion of *children under one year of age* who have received at least one dose of *measles-containing vaccine.* In our application it is the proportion of children of age 0 or 1 at the time of interview whose vaccination card reports a measles vaccination or whose mother recalled the child being given an injection in the upper arm to prevent measles.

Proportion of births attended by skilled health personnel. It is the proportion of total *live births* that are attended by a *skilled birth attendant* trained in providing life-saving obstetric care. In our application it is the proportion of deliveries assisted either by doctor, clinical officer, nurse, midwife or community health worker for all children of aged 0–2 at the time of the interview.

Contraceptive prevalence rate. It is the percentage of *women of reproductive age* (15–49) who are currently using, or whose sexual partner is currently using, at least one contraceptive method. In our application it is the proportion of women aged 15–49 who report using any contraceptive method at the time of the interview.

Antenatal care coverage. It is the percentage of women aged 15–49 with a *live birth* in a given time period that received *antenatal* care provided by *skilled health personnel* at least once during their pregnancy. In our application it is the percentage of women aged 15–49 with a *live birth* that received at least one antenatal visit by either doctor, clinical officer, nurse, midwife or community health worker for children who are aged 0–2 at the time of the interview.

Proportion of population aged 15–24 with comprehensive correct knowledge about HIV/AIDS. The percentage of the population aged 15–24 that has a *comprehensive correct knowledge of Human immunodeficiency virus*. In our application it is the proportion of population aged 15–49 that answered correctly 8 (yes/no) questions about obvious causes of HIV infection transmission.

Proportion of children under-5 sleeping under insecticide-treated bednets. This is the proportion of children aged 0–59 months who slept under an *insecticide-treated mosquito net* the night prior to the survey.

Proportion of the population using an improved drinking water source. In our application it is the percentage of households with access to any of the following sources of drinking water: piped into welling, yard or plot; public tap; tube well and borehole; protected dug well; protected spring; bottles; and sachet water.

Proportion of the population using an improved sanitation facility. In our application it is the percentage of households with access to any of the following improved toilet facilities: flush to piped sewer system; flush to septic tank; flush to pit (latrine); ventilated improved pit latrine; and pit latrine with slab.

Fixed telephone subscriptions for 100 inhabitants. This is the sum of the active number of analogue fixed-telephone lines, voice-over-IP subscriptions, fixed wireless local loop subscriptions, integrated services digital network voice-channel equivalents and fixed public payphones. In our application it is the percentage households reporting having a landline in the home.

Mobile cellular subscriptions for 100 inhabitants. The number of *mobile-cellular telephone subscriptions* per 100 population. In our application it is the percentage of adults aged 15–49 reporting having used a mobile phone.

Appendix A2. Methodology for adjusting for baseline characteristics

This appendix provides an example of the methodology outlined in Section 4 of the report for adjusting the estimation of difference-in-difference effects for the differences in baseline characteristics between the project and the comparison groups. We use the estimation of the effects on per-adult equivalent expenditure and we report the stata commands and their output.

Step 1 (*match.do* dofile) We fit a logit model to the baseline data in which the dependent variable is equal to 1 if the observation is in the MV area and 0 otherwise. We select the following basic strong determinants of the outcomes and project participation: household size, age of head of household, education (in years) of head of household, size of cultivated land, monetary value of household wealth (livestock plus durable assets and productive assets). In addition, we select a number of potential additional covariates that are included to the model stepwise provided they achieve a level of statistical significance equivalent to a P-value below 15%. The results are shown in the output below. The additional variables are: closeness of household (not having relevant ties with other households); polygamous households; female-headed households; having at least one household member migrated for work; having a member sending remittances; not having access to protected water; distance to nearest source of drinking water; not having access to a protected toilet; running a community service business; running a trade business; running a small food business; running any other business; being affected by a drought in the last three years; walls made of mud; floor made of earth; roof made of metal; farmer household; main crop is maize; main crop is millet; main crop is rice; main crop is groundnut; number of months food insecure; having bank savings, being member of *susu*.

```
stepwise, lockterm1 pe(0.15) lr: logit mv ($basic) $additional

Logistic regression                               Number of obs   =       2172

                                                  LR chi2(19)     =     159.74

                                                  Prob > chi2     =     0.0000

Log likelihood = -1293.4521                       Pseudo R2       =     0.0582

------------------------------------------------------------------------------

          mv |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]

-------------+----------------------------------------------------------------

      hhsize |  -.0067876    .013581    -0.50   0.617    -.0334059    .0198306

       hhage |   .0037541   .0032612     1.15   0.250    -.0026378    .0101459

       hhedu |   .0004197   .0171566     0.02   0.980    -.0332067    .0340461

       cland |    .053392    .021823     2.45   0.014     .0106198    .0961643

      wealth |   7.30e-06   .0000101     0.73   0.468    -.0000124     .000027

foodmonthins |  -.1233338   .0303949    -4.06   0.000    -.1829066   -.0637609

   remittant |   .9274683   .2811645     3.30   0.001      .376396    1.478541

    closedhh |   .3450395   .1187084     2.91   0.004     .1123754    .5777036

     flshock |   .6481866   .1459526     4.44   0.000     .3621248    .9342484

      millet |   .4167388   .1133762     3.68   0.000     .1945256    .6389521

        rice |   -.476676   .1086372    -4.39   0.000     -.689601   -.2637509

     drshock |  -.5096336   .1462718    -3.48   0.000    -.7963211   -.2229461
```

```
       gdnut |   .2124261   .1006725     2.11   0.035     .0151117    .4097406
     occfarm |  -.3924285   .1678899    -2.34   0.019    -.7214867   -.0633703
       wdist |   .0025793   .0013494     1.91   0.056    -.0000655     .005224
        bank |   .3832772   .1485601     2.58   0.010     .0921047    .6744498
   metalroof |  -.2654728   .1061399    -2.50   0.012    -.4735032   -.0574425
       maize |     .26521   .1224143     2.17   0.030     .0252823    .5051376
     workmig |   .2376481   .1548082     1.54   0.125    -.0657705    .5410666
       _cons |  -.9903003    .299166    -3.31   0.001    -1.576655   -.4039457
------------------------------------------------------------------------------
```

Next we calculate the squares of all continuous variables selected in the previous model and their interactions with the Builsa dummy variable. We then include squares and interactions stepwise to the previous model specification using a cut-off of significance level equivalent to a P-value of 5%. We then use the coefficient estimates to calculate the propensity score.

```
stepwise, lockterm1 pe(0.05) lr: logit mv ($allvar) *sq bu*

Logistic regression                               Number of obs   =       2172
                                                  LR chi2(33)     =     374.74
                                                  Prob > chi2     =     0.0000
Log likelihood = -1185.9518                       Pseudo R2       =     0.1364
--------------------------------------------------------------------------------
            mv |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
---------------+----------------------------------------------------------------
        hhsize |   .1073372    .041134     2.61   0.009     .0267161    .1879583
         hhage |   .0033048   .0034847     0.95   0.343     -.003525    .0101346
         hhedu |    .261938   .0590696     4.43   0.000     .1461637    .3777122
         cland |    .003488   .0260075     0.13   0.893    -.0474858    .0544618
        wealth |   .0000472   .0000235     2.01   0.044     1.25e-06    .0000932
  foodmonthins |   .1421172   .0824897     1.72   0.085    -.0195596    .3037941
     remittant |   .8214618   .2903768     2.83   0.005     .2523338     1.39059
      closedhh |  -.2074239   .1831543    -1.13   0.257    -.5663999     .151552
       flshock |  -2.177545   .5490342    -3.97   0.000    -3.253632   -1.101458
        millet |    .457753   .1430661     3.20   0.001     .1773487    .7381573
          rice |   .1799297   .1798153     1.00   0.317    -.1725018    .5323612
       drshock |   -1.61066   .2104757    -7.65   0.000    -2.023185   -1.198135
         gdnut |   .2408985   .1089791     2.21   0.027     .0273034    .4544937
       occfarm |  -.4591672   .1801756    -2.55   0.011    -.8123048   -.1060295
         wdist |   .0002876   .0019346     0.15   0.882    -.0035042    .0040793
          bank |   .4424745   .1580461     2.80   0.005     .1327099     .752239
     metalroof |   .0458386   .1460495     0.31   0.754    -.2404131    .3320902
         maize |   .3191898   .1441804     2.21   0.027     .0366014    .6017783
```

```
       workmig |   .2994539   .1622955     1.85   0.065    -.0186395    .6175472
     budrshock |   2.007499   .2749536     7.30   0.000       1.4686    2.546398
        burice |  -1.348602   .2396051    -5.63   0.000    -1.818219   -.8789844
     flshocksq |   2.548018   .4920406     5.18   0.000     1.583636      3.5124
bufoodmonthins |  -.2373618   .0667689    -3.55   0.000    -.3682265   -.1064971
    buclosedhh |   1.203908    .254339     4.73   0.000     .7054129    1.702403
       buhhedu |  -.1441534   .0389876    -3.70   0.000    -.2205677   -.0677391
       hhedusq |  -.0194806   .0050666    -3.84   0.000    -.0294109   -.0095502
      hhsizesq |  -.0060358   .0020152    -3.00   0.003    -.0099855   -.0020862
       bucland |    .201143   .0500789     4.02   0.000     .1029901    .2992958
   bumetalroof |  -.7120687   .2259126    -3.15   0.002    -1.154849   -.2692881
      bumillet |  -.7811166   .2507101    -3.12   0.002    -1.272499   -.2897337
foodmonthinssq |  -.0334938   .0144993    -2.31   0.021     -.061912   -.0050757
       buwdist |   .0058099    .002837     2.05   0.041     .0002495    .0113703
      wealthsq |  -1.49e-09   7.96e-10    -1.87   0.061    -3.05e-09    6.96e-11
         _cons |  -.6830567   .3709084    -1.84   0.066    -1.410024    .0439105
------------------------------------------------------------------------------
```

<u>Step 2</u> (*blocks.do* dofile) We build sub-classification blocks using the algorithm described in section 13.5 of IR and we conduct three tests to assess the validity of the estimated propensity score based on its ability to balance the distribution of the covariates. The first test, section 13.7.1, calculates a Z-value for the statistical significance of the difference of each covariate between project and comparator groups across the blocks. Values are compared to a normal distribution and several large values are a sign of poor balance. None of the reported Z-values is large in our test results.

```
foreach var of varlist $allvar {
  2.         test1 `var'
  3.         }
variable   hhsize
standard t-test   -.83046409
mean difference   -.12865718
Z-value      -.51729939
variable   hhage
standard t-test   -1.5016607
mean difference   -.05588301
Z-value      -.05401946
variable   hhedu
standard t-test   -.97967896
mean difference   -.05554073
Z-value      -.29892628
variable   cland
```

```
standard t-test   -2.6283258

mean difference   -.07612915

Z-value      -.48160558

variable   wealth

standard t-test   -1.7786928

mean difference   -64.016695

Z-value      -.20598334

variable   foodmonthins

standard t-test   3.6083433

mean difference   .02855267

Z-value      .27861044

variable   remittant

standard t-test   -4.1895905

mean difference   .00040176

Z-value      .03069188

variable   closedhh

standard t-test   -3.5636114

mean difference   .00001149

Z-value      .00043862

variable   flshock

standard t-test   -3.2481887

mean difference   -.00427702

Z-value      -.18959147

variable   millet

standard t-test   -2.8163175

mean difference   .00442178

Z-value      .1448278

variable   rice

standard t-test   3.9173785

mean difference   .01264646

Z-value      .42419928

variable   drshock

standard t-test   3.020909

mean difference   -.00756165

Z-value      -.33499847

variable   gdnut

standard t-test   -1.9500163

mean difference   .01407473

Z-value      .43273945
```

```
variable   occfarm

standard t-test   2.5184834

mean difference   .0006588

Z-value      .03404279

variable   wdist

standard t-test   -1.4920189

mean difference   -.08454801

Z-value      -.03717524

variable   bank

standard t-test   -3.1279577

mean difference   -.00708697

Z-value      -.32357298

variable   metalroof

standard t-test   .50191888

mean difference   -.00625899

Z-value      -.19671181

variable   maize

standard t-test   -2.0820942

mean difference   -.00895753

Z-value      -.31352926

variable   workmig

standard t-test   -3.5892526

mean difference   -.00705841

Z-value      -.31650566
```

The second test assesses the overall balance by calculating F-statistics across blocks for each variable. We find only two statistically significant differences: remittances and drought shocks.

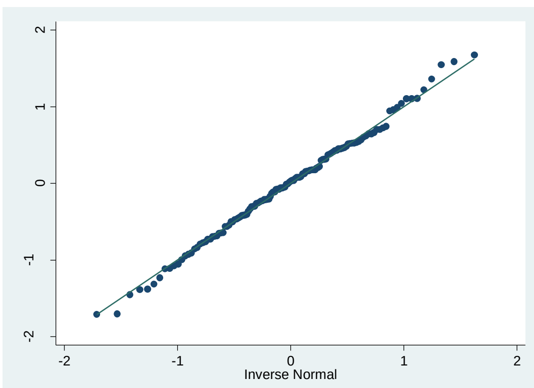
```
foreach var of varlist $allvar {

          test2 `var'

       }

variable   hhsize

F-test slope dummies  1.4198095

P-value  .19274689

variable   hhage

F-test slope dummies  .79524567

P-value  .59123911

variable   hhedu

F-test slope dummies  1.1121703

P-value  .35238857
```

```
variable   cland

F-test slope dummies  1.0256716

P-value  .41078944

variable   wealth

F-test slope dummies  .37346713

P-value  .91813158

variable   foodmonthins

F-test slope dummies  .20007216

P-value  .98551902

variable   remittant

F-test slope dummies  2.4359048

P-value  .01737878

variable   closedhh

F-test slope dummies  .65785879

P-value  .70798579

variable   flshock

F-test slope dummies  1.0207149

P-value  .41431107

variable   millet

F-test slope dummies  .27112307

P-value  .9652049

variable   rice

F-test slope dummies  .50259941

P-value  .83316277

variable   drshock

F-test slope dummies  3.2702407

P-value  .00185055

variable   gdnut

F-test slope dummies  .75633111

P-value  .62413967

variable   occfarm

F-test slope dummies  .26694386

P-value  .96668387

variable   wdist

F-test slope dummies  1.3547133

P-value  .22046158

variable   bank

F-test slope dummies  .43966039

P-value  .87759643
```

```
variable   metalroof

F-test slope dummies  .31166315

P-value  .94901459

variable   maize

F-test slope dummies  .66531628

P-value  .70166676

variable   workmig

F-test slope dummies  .9298393

P-value  .48199678
```

The third test assesses the statistical significance for each covariate and for each block, which in our case consists of 18*7=126 different tests, which we omit for reasons of space. We find however only two statistically significant differences. Finally, we plot the ordered Z-values against the corresponding quantiles of the normal distribution using a QQ plot. The plot follows the 45 degrees line quite closely showing that the distribution is very close to normal.

Figure A2.1 Statistical significant for each covariate



<u>Step 3</u> (*balance.do* dofile) We run three balance checks: normalised difference, log ratio of standard deviations and the proportion of treatment and control observations outside the 95% overlap (section 1.4.2 of IR).

```
foreach var of varlist $allvar pscore linpscore {

  2.         di in white "`var'"

  3.         stdiff `var'

  4.         logratio `var'

  5.         outoverlap `var'

  6.         }

hhsize

normalised difference   .0495227

log ratio of st. deviations   -.09575668

control proportion outside alpha tails   .0568104

treatment proportion outside alpha tails   .03094233
```

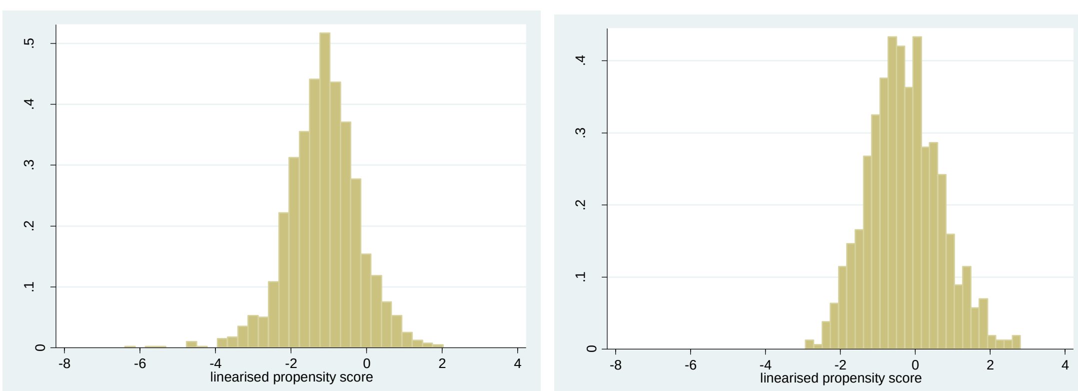
```
hhage

normalised difference   .06617955

log ratio of st. deviations   .01250573

control proportion outside alpha tails   .03353867

treatment proportion outside alpha tails   .05203938

hhedu

normalised difference   .03159934

log ratio of st. deviations   -.04011401

control proportion outside alpha tails   .03216975

treatment proportion outside alpha tails   .00843882

cland

normalised difference   .14148471

log ratio of st. deviations   .05954576

control proportion outside alpha tails   .01848049

treatment proportion outside alpha tails   .02953586

wealth

normalised difference   .0759457

log ratio of st. deviations   -.06378651

control proportion outside alpha tails   .06570842

treatment proportion outside alpha tails   .04360056

foodmonthins

normalised difference   -.23458478

log ratio of st. deviations   -.1543095

control proportion outside alpha tails   .03080082

treatment proportion outside alpha tails   .00140647

remittant

normalised difference   .19797318

log ratio of st. deviations   .50862912

control proportion outside alpha tails   0

treatment proportion outside alpha tails   .05907173

closedhh

normalised difference   .18681061

log ratio of st. deviations   .13981978

control proportion outside alpha tails   0

treatment proportion outside alpha tails   0

flshock

normalised difference   .14635327

log ratio of st. deviations   .0462063

control proportion outside alpha tails   0
```

```
treatment proportion outside alpha tails   0
millet
normalised difference   .11813158
log ratio of st. deviations   -.04202342
control proportion outside alpha tails   0
treatment proportion outside alpha tails   0
rice
normalised difference   -.20584332
log ratio of st. deviations   -.07712222
control proportion outside alpha tails   0
treatment proportion outside alpha tails   0
drshock
normalised difference   -.13708271
log ratio of st. deviations   .13393999
control proportion outside alpha tails   0
treatment proportion outside alpha tails   0
gdnut
normalised difference   .10177533
log ratio of st. deviations   -.00783388
control proportion outside alpha tails   0
treatment proportion outside alpha tails   0
occfarm
normalised difference   -.11508304
log ratio of st. deviations   .16031929
control proportion outside alpha tails   0
treatment proportion outside alpha tails   0
wdist
normalised difference   .08230928
log ratio of st. deviations   .27322462
control proportion outside alpha tails   .01163587
treatment proportion outside alpha tails   .02672293
bank
normalised difference   .14413973
log ratio of st. deviations   .15907652
control proportion outside alpha tails   0
treatment proportion outside alpha tails   0
metalroof
normalised difference   -.03659665
log ratio of st. deviations   -.0093604
```

```
control proportion outside alpha tails   0

treatment proportion outside alpha tails   0

maize

normalised difference   .12774867

log ratio of st. deviations   -.06925947

control proportion outside alpha tails   0

treatment proportion outside alpha tails   0

workmig

normalised difference   .18051674

log ratio of st. deviations   .19331208

control proportion outside alpha tails   0

treatment proportion outside alpha tails   0

pscore

normalised difference   .91493009

log ratio of st. deviations   .23877869

control proportion outside alpha tails   .13826146

treatment proportion outside alpha tails   .1533052
```

Histograms of the linearised propensity scores for the project (left) and comparison group (right).

Figure A2.2 Histograms of linerised propensity scores



We calculate the size of the discrepancy of the estimated propensity scores (section 14.5 of IR), which we find to be very small.

discrepancy

proportion of control units with good matches (0.1 of lin pscore) .96988364

proportion of project units with good matches (0.1 of lin pscore) .99015471

(*trim.do* dofile) We then trim the data using the algorithm described by IR in section 16.4 (equation 16.10). The routine removes 170 observations because they lie outside the region of overlap.

findgamma

difference .00021061

gamma 11.096

alpha left .10015324

alpha right .89984676

(163 real changes made, 163 to missing)

(7 real changes made, 7 to missing)

The reduced sample of observations is the used to re-estimate the propensity score using again the procedure outlined above.

```
stepwise, lockterm1 pe(0.15) lr: logit mv ($basic) $additional

Logistic regression                              Number of obs   =       2002

                                                 LR chi2(17)     =     107.13

                                                 Prob > chi2     =     0.0000

Log likelihood = -1237.7911                      Pseudo R2       =     0.0415

------------------------------------------------------------------------------

          mv |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]

-------------+----------------------------------------------------------------

      hhsize |  -.0012914   .0143693    -0.09   0.928    -.0294547     .026872

       hhage |    .001883   .0033215     0.57   0.571     -.004627    .0083929

       hhedu |   .0168466   .0187223     0.90   0.368    -.0198484    .0535417

       cland |   .0542572   .0218799     2.48   0.013     .0113735     .097141

      wealth |   6.73e-06    .000011     0.61   0.540    -.0000148    .0000283

   remittant |   1.024645   .2576368     3.98   0.000     .5196859    1.529604

      millet |   .3887258   .1116512     3.48   0.000     .1698934    .6075582

        rice |  -.3986168   .1106784    -3.60   0.000    -.6155424   -.1816911

     drshock |  -.5306539   .1480686    -3.58   0.000    -.8208631   -.2404447

     flshock |   .5777997   .1468252     3.94   0.000     .2900275    .8655718

    closedhh |   .3106424   .1191813     2.61   0.009     .0770513    .5442335

foodmonthins |  -.0889394   .0322742    -2.76   0.006    -.1521956   -.0256832

     occfarm |  -.3932619   .1698307    -2.32   0.021     -.726124   -.0603998

        bank |   .3250291   .1495618     2.17   0.030     .0318934    .6181648

   metalroof |  -.1869933   .1089376    -1.72   0.086    -.4005071    .0265206

       wdist |   .0022498   .0013711     1.64   0.101    -.0004376    .0049372

       gdnut |   .1668555   .1021486     1.63   0.102    -.0333521    .3670632

       _cons |  -.6849816   .2927642    -2.34   0.019    -1.258789   -.1111743

------------------------------------------------------------------------------

. stepwise, lockterm1 pe(0.05) lr: logit mv ($allvar) *sq bu*
```

```
Logistic regression                           Number of obs   =       2002

                                              LR chi2(32)     =     288.96

                                              Prob > chi2     =     0.0000

Log likelihood = -1146.8775                   Pseudo R2       =     0.1119


------------------------------------------------------------------------------

           mv |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]

---------------+--------------------------------------------------------------

       hhsize |   .0816777   .0431949     1.89   0.059    -.0029828    .1663382

        hhage |   .0022122   .0035422     0.62   0.532    -.0047303    .0091548

        hhedu |   .2853016   .0638515     4.47   0.000      .160155    .4104482

        cland |  -.0103333     .02766    -0.37   0.709    -.0645459    .0438792

       wealth |   .0000866   .0000277     3.13   0.002     .0000323    .0001408

    remittant |   1.018835   .2668577     3.82   0.000     .4958037    1.541867

       millet |   .2906636   .1305449     2.23   0.026     .0348002    .5465269

         rice |   .1798286   .1816545     0.99   0.322    -.1762078    .5358649

      drshock |  -1.690124    .225017    -7.51   0.000    -2.131149   -1.249098

      flshock |    -2.4137   .5594582    -4.31   0.000    -3.510218   -1.317182

     closedhh |  -.1798306   .1855283    -0.97   0.332    -.5434595    .1837982

  foodmonthins |   .1017443   .0869424     1.17   0.242    -.0686597    .2721484

      occfarm |  -.4953135   .1818301    -2.72   0.006    -.8516941    -.138933

         bank |   .4468037   .1601084     2.79   0.005     .1329971    .7606104

    metalroof |  -.0022758    .148482    -0.02   0.988    -.2932951    .2887435

        wdist |  -.0000657   .0019726    -0.03   0.973    -.0039319    .0038005

        gdnut |    .241563    .110055     2.19   0.028     .0258593    .4572668

   buclosedhh |   1.211973   .2590827     4.68   0.000     .7041805    1.719766

    flshocksq |   2.763417   .5027162     5.50   0.000     1.778111    3.748722

       burice |  -1.269488   .2461548    -5.16   0.000    -1.751942   -.7870331

    budrshock |   2.154957   .3129302     6.89   0.000     1.541625    2.768289

       builsa |  -1.255219   .3452311    -3.64   0.000    -1.931859   -.5785784

      bucland |   .2566972   .0554878     4.63   0.000     .1479431    .3654512

     hhedusq |  -.0229961   .0056356    -4.08   0.000    -.0340418   -.0119504

      buhhedu |  -.1326067   .0441543    -3.00   0.003    -.2191476   -.0460658

    wealthsq |  -2.36e-09   9.25e-10    -2.55   0.011    -4.17e-09   -5.47e-10

     buwealth |   -.000058   .0000252    -2.30   0.021    -.0001074   -8.59e-06

foodmonthinssq |  -.0330801   .0160521    -2.06   0.039    -.0645417   -.0016185

      buwdist |   .0066624   .0029421     2.26   0.024      .000896    .0124288

     hhsizesq |  -.0045643   .0021575    -2.12   0.034    -.0087929   -.0003357

   bumetalroof |  -.5422138   .2362127    -2.30   0.022    -1.005182   -.0792454
```

```
bufoodmonthins |  -.1612239   .0725831    -2.22   0.026    -.3034842   -.0189636
         _cons |   .1134293   .3928537     0.29   0.773    -.6565497    .8834083
------------------------------------------------------------------------------
```

Appendix A3. Monetary poverty

Imputation of expenditure items

We made some additional imputations in the calculation of expenditure figures that we had not made at the baseline and first follow-up. Following GSS practice we estimate housing expenditure by imputing house rents. The methodology employed by GSS is the following. First, a subset of households paying rents is obtained from the data. Rents paid in cash and kind are added up, the variable is transformed in logarithms and outliers (outside 3sd from the logmean) are removed. Second, house rents are regressed on a number of house characteristics such as, floor, wall and toilet type. Third, the regression coefficients are used to make out of sample predictions for those households not paying any rent.

Our dataset, mainly composed of poor rural households, contains very few households paying rents so that rent regressions within the MVP sample are not possible. Therefore we estimated rent regressions using the GLSS6 data of 2012/2013 and made out of sample predictions to our dataset. We included in the regressions the following house characteristics that are common to the MVP and GLSS datasets and that are strong determinants of rent values: number of bedrooms, water piped in the home, electricity, flush toilet in the home, mud walls, cement/brick walls, mud floors and regional variables for the 10 administrative regions by urban/rural.

Nominal rents calculated in this way easily apply to the Round 2 data that were collected with reference to the same agricultural year of GLSS6. However, house rent predictions at the baseline, midterm and following rounds need to be adjusted for inflation. Regional deflators are applied separately to the Builsa and West Mamprusi data to express rents in nominal values corresponding to the period in which the data were collected.

Finally, note that the Round 2 data do not contain information on housing conditions. Therefore we decided to calculate rents for Round 2 using housing conditions at the baseline. In this way, house rents and total expenditure at Round 2 can be partly underestimated if housing conditions have improved between the baseline and Round 2.

Equivalence scales

Equivalence scales are used to determine poverty status of households. We use the same equivalence scale used by GSS. These scales are reported on page 75 of GSS (2014) Poverty Profile in Ghana (2005-2013) and the source is: Recommended Dietary Allowances (1989), Washington, DC: National Academy Press. These equivalence scales are rather old and different from those recommended by WHO/FAO (Human Energy Requirements: Report of a joint FAO/WHO/UNU expert consultation, (2004) Rome: United Nations University). The latter scales give higher weight to adults and lower weight to children in comparison to older scales. Hence households with many children have fewer adult equivalents using the newer scale thus reducing some of the correlation between poverty and household size.

Inflation adjustments

Since the Ghanaian poverty line is calculated at Accra prices of January 2013 and since January 2013 is the midpoint of reported expenditure of our second round of data, we decided to calculate real expenditures across survey rounds in relation to prices prevailing in the agricultural year 2012/2013. Our consumer price indices are therefore set to 1.0 at January 2013. The price deflators are then calculated using monthly price series downloaded from the GSS website (some of the points in the series are missing and were simply interpolated). We calculated the deflators separately for Builsa and West Mamprusi districts.

We used the Northern Region series for prices in West Mamprusi and the Upper East series for prices in Builsa. Since the data were collected with a three-month gap between project and control areas at the baseline, we also calculated deflators separately for the project and the control groups at the baseline. Price series are disaggregated into food and non-food. We calculated a combined price index using food expenditure shares weights obtained from the MVP data (separately for Builsa and West Mamprusi) at the second round. Food and non-food items have different price dynamics and the food expenditure share in our sample is quite large (around 65%). Our weighted combined index better reflects how households in our sample are affected by price changes than the combined index provided by GSS. The indices and summary calculations can be found in the excel file CPI in the Round 3 folder.

Poverty lines

We used the poverty line per-adult equivalent calculated by the GSS: an overall poverty line of 1314 Cedis and a food poverty line of 792.05 Cedis. These are the poverty lines in Accra at January 2013 prices. Household expenditure needs to be adjusted for regional differences in prices (the same good costs differently in Accra and Builsa) and temporal differences in prices (the same good has a different price if the interview takes place in January or in June).

Regional differences are corrected using the deflator calculated by the GSS and reported in the GLSS6 aggregate poverty data. We deflate the poverty line by the regional rural deflator of the Northern region in the case of West Mamprusi (0.9825) and by the regional rural deflator of the Upper East in the case of Builsa (0.9306). Life is cheaper in the North and this is reflected in the regional deflators, however, the difference does not appear to be very large. For example, the poverty lines at January 2013 prices are 1,412 and 851 for Builsa and 1,337 and 806 for West Mamprusi.

Time differences are corrected using the combined CPI regional indices described above. In the case of the Round 2 data no adjustment is made because the poverty line was set at January 2013 prices which is exactly the mid-point of the MVP data collection in 2012/2013. The MVP data were collected over periods of two to three months, however for most items (food and low frequency non-food) the recall period is 12 months so that some 80% of expenditures are reported in prices over the previous 12 months, not the prices at the time of the interview. Rather than adjusting poverty lines for inflation, we adjust expenditures at baseline, midterm and following rounds. The deflators to calculate expenditures at 2013 prices are reported below. Note that there are different deflators for the project and control areas at the baseline because the data were collected at different times of the year in the two areas. In the following years the data were collected simultaneously in the project and control areas so that the same deflator can be applied within the same region.

Table A3.1 Price deflators

	Builsa project	Builsa control	West Mamprusi project	West Mamprusi control
2011/2012	0.969	0.978	0.943	0.958
2012/2013	1.000	1.000	1.000	1.000
2013/2014	1.090	1.090	1.141	1.141

Appendix A4. Quality of Mortality Data

Data on mothers and full birth histories

Mortality data were collected from full birth history interviews in the adult survey. In the adult survey all women of reproductive age (15–49) that had been listed as household members were interviewed. The follow-up survey of 2014 reported a women's response rate of 94%. Table A4.1 reports the number of households and adults interviewed. More households and women were interviewed at the follow-up.

Table A4.1 Sample sizes of the baseline and midterm adult surveys

	Baseline	Midterm
Households interviewed	2,013	2,120
Male adults (15–49) interviewed	1,656	1,836
Female adults (15–49) interviewed	2,894	3,242
Female who ever gave birth	2,187	2,355
Birth histories	9,536	10,283

The sample size of the MV survey is smaller than the standard nationally representative sample used by DHS in Ghana (Table A4.2), particularly considering that the MV project sample is just one-third of the total sample (about 3,000 birth histories). It is common however for the DHS to report mortality rates by region, of which there are 10 in Ghana. From this point of view the MV sample is not smaller than the standard DHS sample.

Table A4.2 Sample sizes of full birth histories in DHS and MV datasets

Survey year	Birth histories
DHS 1988	14,216
DHS 1993	13,298
DHS 1998	13,188
DHS 2003	15,086
DHS 2008	11,888
MICS 2011	31,145
MVP baseline 2012	9,536
MVP follow-up 2014	10,283

Many of the women interviewed at baseline were re-interviewed at the first follow-up. However, the survey employed the same household codes across surveys but not the same individual codes. Hence, it is not possible to follow a panel of women across the two surveys.

Missing values

A first group of missing women consists of women who were not enumerated in the household roster either because they died or migrated before the interview. The birth histories relating to these mothers should contribute to the definition of mortality rates. To the extent that mother mortality and migration are correlated to child mortality, the observed sample is biased and does not correctly represent mortality rates in the population. A second group of missing women consists of those women that were enumerated in the household roster but nevertheless not interviewed during the adult interviews. This is another potential source of selection bias. This source can be checked by looking at differences in known determinants of mortality between women interviewed and not interviewed. For example, child mortality is correlated with mother's age and therefore a difference in age between the two groups will result in a biased estimate of mortality rates in the population.

Though the goal of the adult survey was to interview all male and female adults of reproductive age (15–49), this was not always possible. We matched the adult data to the household roster data to see what completion rate was achieved by sex. This is reported in the table below. Note also that there is a fraction of women interviewed by the adult survey but not reported in the household roster because apparently they were not classified as household members at the time of the household interview. At the baseline there were 63 female and 30 male respondents of this type, while at the midterm there were 31 female and 21 male of this type.

Table A4.3 Percentage of household members (age 15–49) interviewed in the adult survey

	Baseline		Midterm	
	Male	Female	Male	Female
Not interviewed	52.8	22.4	51.8	23.0
Interviewed	47.3	77.6	48.2	77.0

Women's interviews are correlated with age and presumably with other variables that are correlated with child mortality. In particular, completion rate increase with age, it reaches a peak at the age group 30–44 and then decreases again. There are differences in the age distribution of mothers across the two surveys too. However, we do not observe the phenomenon observed in the DHS data whereby mothers in the age groups 15–19 and 45–49 are underrepresented, presumably artificially generated by enumerators with the purpose of avoiding the full birth history interview. In the MV study the adult and household surveys are conducted separately so that incentives and opportunities for this are lower.

Table A4.4 Age distribution of eligible and interviewed women

Age group	Eligible women	% interviewed	Eligible women	% interviewed
10–14	890	0.3	1,057	1.5
15–19	773	71.9	969	71.6
20–24	650	78.0	779	68.4
25–29	609	78.8	668	75.5
30–34	453	82.6	513	83.0
35–39	453	81.0	462	84.9
40–44	364	77.2	393	87.8
45–49	301	76.4	341	83.0
50–54	252	5.2	292	2.4

Dates of births and deaths are sometimes missing particularly in the midterm survey. About 50% of children whose date of birth is missing are reported dead which results in an underestimation of mortality particularly for the midterm survey for which the phenomenon is more frequent. Since the age at death is rarely missing, a method for the imputation of age at birth is advisable particularly when the month of birth is missing but the year of birth is available. We adopted a simple fix consisting of using a midpoint imputation: children born in a given year are reported as being born in June when the month is missing. We then used a random month imputation to avoid a clustering of child deaths in mid calendar year. The month of birth is imputed drawing from a uniform distribution. The procedure does not generate inadmissible months of birth as there are no children born in recent years who could be reported as dying after the survey month because of the imputation procedure (most recent deaths occurred within few months from birth) and no month of birth is missing for children born in the year of the survey (2012 at baseline and 2014 at the midterm).

Table A4.5 Missing dates of birth and death

	Missing date of birth	Of which are dead	Missing age at death
Baseline	0.14 (14)	57.1 (8)	0.02 (2)
Midterm	3.81 (392)	46.9 (194)	0.00 (0)

Sex ratios should be around 100 or slightly above so that sex ratios decreasing over time are a sign of selective error in reporting. Sex ratios vary widely from year to year but not obvious pattern emerges in either survey. Birth ratios calculated as in Hill (2013) also highlights errors in reporting, particularly because children under-5 require the collection of much additional data in the remaining of the questionnaire thus creating an incentive for enumerators for putting age of birth back in time. This patterns does appear in the data were the number of births five years before the survey is considerably less than at six years before the survey a pattern similarly observed in DHS surveys in Ghana.

Table A4.6 Sex ratios and births ratios by calendar year

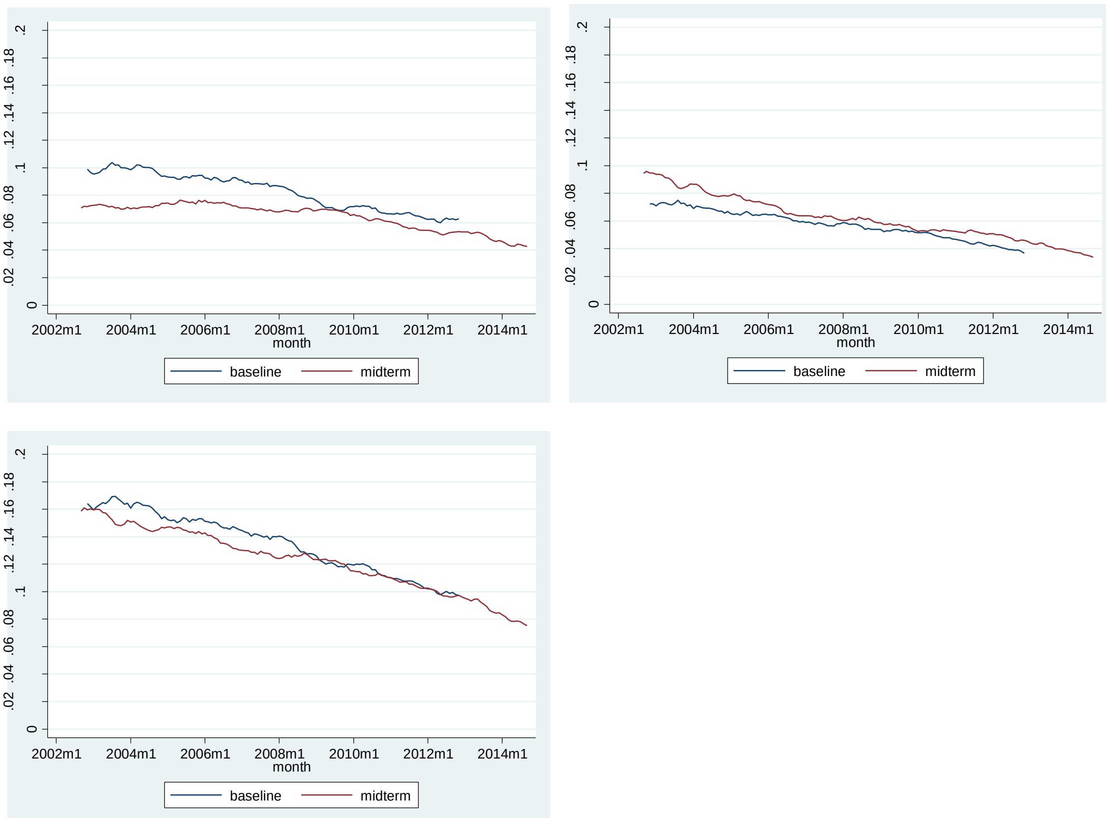
Calendar year	Baseline		Midterm	
	Sex ratio	Birth ratio	Sex ratio	Birth ratio
1993	0.83	0.73	0.88	0.80
1994	0.99	1.35	1.00	1.15
1995	1.06	0.86	1.06	0.97
1996	0.99	0.93	1.07	1.10
1997	0.95	1.10	1.05	0.87
1998	1.03	1.03	1.02	1.07
1999	0.99	0.88	1.07	0.99
2000	0.79	1.13	0.94	1.07
2001	0.87	0.93	0.93	0.93
2002	0.94	1.03	1.06	0.99
2003	1.02	0.94	0.77	1.01
2004	1.01	1.13	0.96	1.02
2005	0.91	0.85	0.82	1.03
2006	1.00	1.21	1.02	0.97
2007	1.00	0.88	0.94	0.93
2008	1.00	1.01	0.99	1.21
2009	0.96	1.01	1.01	0.88
2010	0.92	0.98	1.03	1.01
2011	0.80	1.30	0.99	1.02
2012	0.99	1.19	0.93	0.97
2013			1.07	1.30
2014			0.86	1.12

Cohort analysis

A powerful method to calculate the accuracy of mortality data is the analysis of overlapping cohorts. The mortality rates calculated in any particular calendar year or month should be very similar for the baseline and midterm survey considering that they are often the same mothers to be interviewed. A perfect match between the two cannot be expected because a large fraction of mothers interviewed has changed between the baseline and the midterm survey and because surveys at the midterm and baseline were at different times of the year.

The figures below show the main patterns. The midterm data appear to underestimate infant mortality and overestimate child mortality in comparison to the baseline data. Overall the under-5 mortality rates are very similar from 2008 (five years before the baseline survey) but diverge a bit before that date, with midterm data underestimating the number of deaths.

Figure A4.1 Infant, child and under-5 mortality at baseline and midterm



References

Angrist, J.D., and J.S. Pischke. 2009. *Mostly Harmless Econometrics*. Princeton: Princeton University Press.

Becker, G. 1991. *A Treatise on the Family*. Cambridge: Harvard University Press.

Cowell, F.A. 2011. *Measuring Inequality*. Oxford: Oxford University Press.

Deaton, A. 1997. *The Analysis of Household Surveys*. Washington, DC: The World Bank.

Fink, G., M. McConnell, and S. Vollmer. 2014. Testing for heterogeneous treatment effects in experimental data: false discovery risks and correction procedures. *Journal of Development Effectiveness* 6 (1):44-57.

Ghana Statistical Service. 2011. *Ghana Multiple Indicator Cluster Survey with an Enhanced Malaria Module and Biomarker*, *Final Report*. Accra Ghana.

Glewwe, P. 2000. Household Roster. In *Designing Household Survey Questionnaires for Developing Countries*, edited by M. Grosh and P. Glewwe. Washington, DC: World Bank.

Grossbard, A. 1980. The Economics of Polygamy. In *Research in Population Economics Vol II*, edited by J. Da Vanzo and J. Simon. Greenwich, CT: JAI-Press.

Guyer, J.I. 1981. Household and community in African studies. *African Studies Review* 24 (2/3):87-137.

Halliday, t. J. (2005). "Mismeasured Housheold Size and its Implication for the Identification of Economies of Scale." Oxford Bulletin of Economics and Statistics 72(2): 246-262.

Hill, K. 2013. Direct estimation of child mortality from birth histories. In *Tools for demographic estimation*, edited by T.A. Moultrie, R.E. Dorrington, A.G. Hill, K. Hill, I.M. Timmaeus and B. Zaba. Paris: International Union for the Scientific Study of Population.

Hill, P. 1986. *Development Economics on Trial: the Anthropological Case for a Prosecution*. New York: Cambridge University Press.

IEG. 2010. *What Can We Learn from Nutrition Impact Evaluations: Lessons from a Review of Interventions to Reduce Child Malnutrition in Developing Countries*. Washington, DC: World Bank.

Imbens, G.W., and J.M. Wooldridge. 2009. Recent Developments in the Econometrics of Impact Evaluation. *Journal of Economic Literature* 47 (1):5-86.

Inbens, G., and D. Rubin. 2015. *Causal Inference for Statistics, Social and Biomedical Sciences*. New York: Cambridge University Press.

Masset, E., L. Haddad, A. Cornelius, and J. Isaza-Castro. 2012. Effectiveness of interventions that aim to improve nutritional status of children: systematic review. *British Medical Journal* 344 (d8222):1-7.

Rosenbaum, Paul R. 1984. The Consequences of Adjustment for a Concomitant Variable That Has Been Affected by the Treatment. *Journal of the Royal Statistical Society. Series A (General)* 147 (5):656-666.

Schochet, Peter Z. 2008. Technical Methods Report: Guidelines for Multiple Testing in Impact Evaluations. *Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education*.

Whitehead, A. 2006. Persistent Poverty in North East Ghana. *Journal of Development Studies* 42 (2):278-300.